SIMPLE ROBUST FIXED LAG SMOOTHING

by

N. D. Le
R. D. Martin
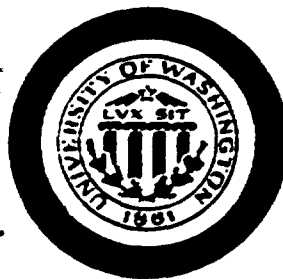
TECHNICAL REPORT No. 149

December 1988

Department of Statistics, GN-22

University of Washington

Seattle, Washington 98195 USA

Simple Robust Fixed Lag Smoothing

With Application To Radar Glint Noise *

N. D. Le   R. D. Martin

Department of Statistics, GN-22
University of Washington
Seattle, WA 98195

December 2, 1988

Abstract

This paper introduces a class of robust lag-k smoothers based on simple low order Markov models for the Gaussian trend-like component of signal plus non-Gaussian noise models. The kth order Markov models are of the kth difference form $\Delta^k x_t = \epsilon_t$ where $\Delta x = x_t - x_{t-1}$ and ϵ_t is a zero-mean white Gaussian noise process with variance σ_ϵ^2. The nominal additive noise is a zero-mean white Gaussian noise sequence with variance σ_o^2, while the actual additive noise is non-Gaussian with an outliers generating distribution, e.g., $(1 - \gamma)N(0, \sigma_o^2) + \gamma H$. This setup is particularly appropriate for radar glint noise. Implementation of the smoothers requires estimation of the two parameters σ_ϵ^2 and σ_o^2. This is accomplished using a robustified maximum likelihood approach. Application to both artificial data sets and to glint noise data reveals that the approach is quite effective. We briefly discuss the choices of lag k for the smoothers and also briefly study the sensitivity of our approach to model mismatch.

1 Introduction

The motivation for this paper is the desire to construct robust fixed lag smoothers which work well on radar glint noise, and which have a simple structure with at most a few parameters to estimate. We achieve this end by both generalizations and specialization of known methods (Martin, 1979; Martin and Thompson, 1982) for constructing robust filters which are insensitive to outliers.

One generalization consists of building on the known robust filter algorithms to obtain robust fixed-lag smoothers. The emphasis here is on fixed-lag smoothing, as opposed to the use of existing robust fixed interval smoothers (e.g., as in Martin, 1979; Martin and Yohai, 1985), because we ultimately wish to use such smoothers in real time applications.

A second generalization is that of estimating the model parameters, including the standard deviation of a nominally Gaussian additive noise model, using a robustified maximum likelihood method.

The specialization is guided by our prior knowledge about radar glint noise. Such noise contains both a low frequency approximately Gaussian trend-like core component, and a broad-band highly non-Gaussian component associated with glint noise "spikes" or outliers. The goal of the robust smoothing is to produce a smooth version of the trend-like core component. Prior experiments (Martin and Thomson, 1982) with robust autoregressive model fitting and robust smoothing indicate that the trend-like component is well modeled by 1st, 2nd or at most 3rd order autoregressive models with parameters on the boundary of nonstationarity. More precisely, state variable models wherein 1st, 2nd or 3rd differences of the state variable are equal to a Gaussian white noise innovations process have appeared to work quite well (correspondingly the state variable is once, twice or thrice "intergrated" white noise). Thus our specialization is to use such simple state variable models to model the trend-like component. Correspondingly, it is not necessary to estimate the autoregressive parameters in the state variable model, only an estimate of the innovations variance σ_{ϵ}^2 for the state process is required.

It is assumed that there is additive white noise v_t which is well modeled by the mixture model

$$F_v = (1 - \gamma)N(0, \sigma_o^2) \ + \ \gamma H \tag{1}$$

with Gaussian central component $N(0, \sigma_o^2)$, and with contamination distribution H. The latter is a heavy-tailed outliers generating distribution which accounts for the glint noise spikes.

Past experience shows that it is not terribly important to estimate the fraction of contamination γ - presumption of any value in the range .1 to .25 will usually do quite well. However, estimates of σ_o^2 is quite important, since σ_o^2 and σ_ϵ^2 determine the degree of smoothing of our smoothers, just as in the case of linear smoothers. Thus there are just two unknown parameters, σ_o^2 and σ_ϵ^2, to be estimated.

There is also a "tuning constant" c which controls the tradeoff between robustness and efficiency for the robust smoother. This constant can be set in advance using the study of Martin and Su (1985) for guidance.

The paper is organized as follows. Section 2 reviews robust filters for the state-variable setup of interest. Section 3 presents robust fixed-lag smoothers. Section 4 discusses the robust estimation of the unknown model parameters σ_o^2 and σ_ϵ^2. Section 5 presents the application of the method to the radar glint noise data. Section 6 discusses the choice of lag. Section 7 considers the issue of sensitivity of the method to model mismatch. Section 8 contains some concluding remarks.

2 The State Variable Setup and Robust Filters

2.1 The State Variable Setup

Suppose that the scalar observations $y_1, ..., y_n$ are generated by the following state-variable system:

$$
\begin{aligned}
x_t &= \Phi x_{t-1} + \epsilon_t \\
y_t &= H x_t + v_t,
\end{aligned}
\tag{2}
$$

where x_t and ϵ_t have dimension p, Φ is a $p \times p$ matrix and H is a $1 \times p$ matrix. It is also assumed that x_t is independent of future ϵ_t, and that ϵ_t, v_t are mutually independent sequences which are individually independent and identically distributed (i.i.d). The innovation process ϵ_t for the state equation is assumed to be a white noise zero mean Gaussian process with covariance matrix Q. The additive noise process v_t is assumed to be zero mean white noise non-Gaussian mixture distribution given by (1).

An example of this system is the pth order autoregression

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + \epsilon_t. \tag{3}$$

This model has the above state-variable form with

$$\Phi = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \tag{4}$$

$$\begin{aligned} x_t^T &= (x_t, x_{t-1}, \ldots, x_{t-p+1}) \\ H &= (1, 0, \ldots, 0) \\ \epsilon_t^T &= (\epsilon_t, 0, \ldots, 0). \end{aligned}$$

The notation to be used throughout is as follows. Denote the first t observations by $Y^t = (y_1, ..., y_t)$. The state prediction density is $f(x_{t+1}|Y^t)$ is the conditional density of x_{t+1} given Y^t. This density is assumed to exist for $t \geq 1$. The observation prediction density is $f(y_{t+1}|Y^t)$. The conditional-mean estimate of x_t given Y^t is written $\hat{x}_t = E(x_t|Y^t)$ and the conditional-mean predictor of x_{t+1} given Y^t is written $\hat{x}_{t+1}^t = E(x_{t+1}|Y^t)$. The conditional-mean estimate of x_t given Y^{t+l} for a fixed constant $l \geq 0$, is written $\hat{x}_t^{t+l} = E(x_t|Y^{t+l})$. This estimate is called a fixed-lag (lag l) smoother. Assuming they exist, these conditional-means are the minimum mean-squared-error linear estimates given Y^t or y^{t+l} (Meditch, 1969).

3

2.2 The pth Order Difference Models

In this subsection, we describe the pth order difference model and its state variable setup. The pth order difference model is defined as

$$\Delta^p x_t = \epsilon_t$$

where

$$\Delta^1 x_t = x_t - x_{t-1}$$
$$\Delta^p x_t = \Delta^{p-1} x_t - \Delta^{p-1} x_{t-1},$$

and ϵ_t has a white noise Gaussian distribution with mean 0 and variance σ_ϵ^2. These models are the nonstationary limiting case of stationary autoregressive models as the coefficients tend to particular points on the boundary of stationarity. For example, the 2nd order ($p = 2$) differences model can be presented as

$$x_t = 2x_{t-1} - x_{t-2} + \epsilon_t$$

which is the same as in (3) with $\phi_1 = 2$ and $\phi_2 = -1$. So the state variable setup for the pth order difference model is similar to (4) with $\phi_1, ..., \phi_p$ appropriately specified.

As we mention in the introduction, the difference models with order $p \leq 3$ are used for our fixed-lag smoothers. We now present the state variable setup for the 1st, 2nd and 3rd order difference models. The Φ, H, ϵ and x for each order are as follows.

Case $p = 1$:

$$\Phi = 1$$
$$x_t = x_t$$
$$H = 1$$
$$\epsilon_t = \epsilon_t.$$

Case $p = 2$:

$$\Phi = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}$$

4

$$x_t^T = (x_t, x_{t-1})$$
$$H = (1, 0)$$
$$\epsilon_t^T = (\epsilon_t, 0).$$

Case $p = 3$:

$$\Phi = \begin{pmatrix} 3 & -3 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

$$x_t^T = (x_t, x_{t-1}, x_{t-2})$$
$$H = (1, 0, 0)$$
$$\epsilon_t^T = (\epsilon_t, 0, 0).$$

The choice of such state variable models for smoothing problems is motivated by the connection between continuous time versions of such models and spline smoothing (see for example, Wecker and Ansley, 1983).

2.3 Masreliez's Filter

We now discuss the methods for estimating the conditional means of x_t given Y_t in the system (2). When ϵ_t and v_t in system (2) are Gaussian, the straight forward computation of $\hat{x}_t = E(x_t|Y^t)$ by any one of a variety of techniques (Jazwinski, 1970) yields the Kalman filter recursion equations. Unfortunately, the explicit and exact calculation of $\hat{x}_t$ in closed form is virtually intractable when v_t is non-Gaussian, except in a few special cases, e.g., as in the case of stable random variable (see Stuck, 1976). However, there is a simplifying assumption, discovered by Masreliez (1975), which allows one to make an exact calculation of $\hat{x}_t$. Masreliez's assumption is that the state predictor density is Gaussian

$$f(x_t|Y^{t-1}) = N(x_t; \hat{x}_t^{t-1}, M_t) \tag{5}$$

where $N(\cdot; \mu, \Sigma)$ denotes the multivariate normal density with mean μ and covariance matrix Σ, and

$$M_t = E\{(x_t - \hat{x}_t^{t-1})(x_t - \hat{x}_t^{t-1})^T|Y^{t-1}\}. \tag{6}$$

is the conditional covariance matrix for the state prediction error. Under this assumption, Masreliez (1975) showed that $\hat{x}_t = E(x_t|Y^t)$, $k \geq 1$ can be obtained by the recursions

$$\hat{x}_t = \hat{x}_t^{t-1} + M_t H^T \Psi_t(y_t) \tag{7}$$

$$M_{t+1} = \Phi P_t \Phi^T + Q \tag{8}$$

$$P_t = M_t - M_t H^T \Psi_t'(y_t) H M_t \tag{9}$$

where

$$\Psi_t(y_t) = -(\frac{\partial}{\partial y_t}) \log f_Y(y_t|Y^{t-1}) \tag{10}$$

is the "score" function for the observation prediction density $f_Y(y_t|Y^{t-1})$, Q is the covariance matrix of ϵ_t, $\hat{x}_t^{t-1} = \Phi \hat{x}_{t-1}$ and

$$\Psi_t'(y_t) = -(\frac{\partial}{\partial y_t}) \Psi_t(y_t). \tag{11}$$

Comments and discussions on the appropriateness of Masreliez's assumption and the validity of the representation (7)-(10) can be found in Martin (1979), and Fraiman and Martin (1988).

2.4 ACM-Type Robust Filter for Autoregressions

Utilizing the Masreliez result, Martin (1979) suggests an approximate conditional mean (ACM) type robust filter described as follows. Let σ_0^2 be the variance of the Gaussian component of the distribution (1) for v_t, and let Ψ be a robustifying psi function as in the robustness literature (see for example, Huber 1981; Hampel et al 1986). Set

$$s_t^2 = m_{1k} + \sigma_0^2$$

where m_{1k} is the 1-1 element of M_t the state prediction error covariance matrix. Then the recursions (7)-(9) are replaced by

$$\hat{x}_t = \Phi \hat{x}_{t-1} + \frac{m_t}{s_t^2} s_t \Psi_c(\frac{r_t}{s_t}) \tag{12}$$

$$M_{t+1} = \Phi P_t \Phi^T + Q \tag{13}$$

$$P_t = M_t - w(\frac{r_t}{s_t}) \frac{m_t m_t^T}{s_t^2} \tag{14}$$

where $\mathbf{m}_t$ is the first column of M_t and r_t is the observation prediction residual

$$r_t = y_t - \hat{x}_t^{t-1}.$$

Two possible choices of w are

$$w(r_t) = \Psi_c'(r_t) \tag{15}$$

by analogy with (11), or

$$w(r_t) = \frac{\Psi_c(r_t)}{r_t}. \tag{16}$$

There are many possible choices for the psi-function in the robust filtering context. These choices are described in Martin (1979, 1981), Martin and Su (1985). The main qualitative characteristics of a good robustifying ψ is that ψ should be bounded and "redescending". For example, Hampel's two-part redescending function defined as

$$\Psi_{HA,c}(y) = \begin{cases} y & |y| \le \alpha c \\ \alpha(c-y)/(\alpha-1) & \alpha c < y \le c \\ -\alpha(c+y)/(1-\alpha) & -c \le y < -\alpha c \\ 0 & |y| > c \end{cases} \tag{17}$$

has these features. Throughout this paper, we use $\Psi_{HA,c}$ with $\alpha c = 1.6$ and $c = 4.0$. These are values suggested by the study of Martin and Su (1985).

We make the choice (16) in this paper since it has a continuity/resistance rationale (see Martin and Su, 1985): the w in (16) is continuous for piecewise linear Ψ_c, whereas the w in (15) is discontinuous.

3 Robust Fixed-Lag Smoothers

3.1 Robust Fixed Lag Smoothers

The above robust filter can be used to obtain a robust fixed lag (lag l, for some $l \le n$) smoother for quite general state equations by augmenting the state equation (see for

example, Anderson and Moore, 1979). In this approach, the state-variable system (2) is augmented as follows:

$$z_t = \Phi_1 z_{t-1} + \epsilon_t^1$$
$$y_t = H_1 z_t + v_t, \tag{18}$$

where

$$\Phi_1 = \begin{pmatrix} \Phi & 0 & \dots & 0 & 0 \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I & 0 \end{pmatrix}$$

$$z_t^T = (x_t, x_{t-p-1}, \dots, x_{t-l})$$
$$H_1 = (H, 0, \dots, 0)$$
$$(\epsilon_t^1)^T = (\epsilon_t^T, 0, \dots, 0).$$

The vectors z_t, ϵ_t^1 have dimension $p(l+1)$. Φ_1 is a $p(l+1) \times p(l+1)$ matrix. H_1 is a $1 \times p(l+1)$ matrix. So for $0 \le i \le l$, the robust fixed lag (i) smoother is just the $(i+1)^{th}$ element of $\hat{z}_t = E(z_t|Y^t)$. The conditional-mean vector $\hat{z}_t$ is obtained by the robust filter.

The state-augmentation approach to fixed lag smoother construction results in state vectors and matrices with dimension proportional to the desired lag. In particular, the dimensions increase by p if the lag is increased by 1. In general, this may become computationally burdensome. However, this problem can be avoided for special case of autoregression type state equations since there is a lower dimensional augmentation in such cases. For the pth order autoregression (3), including the pth

order difference model, with lag $l \geq p$, we can use (17) with

$$
\Phi_1 = \begin{pmatrix}
\phi_1 & \phi_2 & \cdots & \phi_p & 0 & \cdots & 0 & 0 \\
1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0 & \cdots & 0 & 0 \\
0 & 0 & \cdots & 0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 & \cdots & 1 & 0
\end{pmatrix}
\tag{19}
$$

$$
\begin{aligned}
z_t^T &= (x_t, x_{t-1}, \ldots, x_{t-l}) \\
H_1 &= (1, 0, \ldots, 0) \\
(\epsilon_t^1)^T &= (\epsilon_t^T, 0, \ldots, 0).
\end{aligned}
$$

The vectors z_t, H_1, ϵ_t^1 have dimensions $(l + 1)$, and Φ_1 is a $(l + 1) \times (l + 1)$ matrix. With this augmentation method, the dimension increases by units of order 1 if the lag is increased by 1. More detail on reducing dimensionality can be found in Moore (1973), Anderson and Moore (1979).

3.2 The pth Order Difference Model Case

For the pth order difference model, we can obtain the fixed-lag smoothers (lag 1 to lag l) by using the setup in (18). Recall that the pth order difference model is just the pth order autoregressive model with $\phi_1, \ldots, \phi_p$ appropriately chosen. To illustrate the approach more specifically, we present the detail of how to obtain the lag-2 and lag-5 smoothers with the 2nd order difference model. This would hopefully help to explain the method more clearly. Using the setup (18), the 2nd order difference model for lag

9

5 smoother can be presented by

$$\Phi_1 = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$z_t^T = (x_t, x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5})$$
$$H_1 = (1, 0, 0, 0, 0, 0)$$
$$(\epsilon_t^1)^T = (\epsilon_t, 0, 0, 0, 0, 0).$$

Suppose that the parameters σ_o^2, σ_ϵ^2 and the tuning constant c are known, we can obtain the conditional mean of the vector random variable z_t conditioned on the observations $y_1, ..., y_t$ by using the ACM-type filter algorithm described above. The lag-2 smoother, $\hat{x}_{t-2} = E(x_{t-2}|Y^t)$, is then the 3rd component of $\hat{z}_t$. The lag-5 smoother is then the 6th (last) component of $\hat{z}_t$.

However, the parameters σ_o^2 and σ_ϵ^2 are unknown and need to be estimated. In the next section, a method for estimating these parameters with p^{th} order difference model is discussed.

4 Parameter Estimation

4.1 Robust Maximum Likelihood Type Estimation

Suppose that the observations $y_1, ..., y_n$ are generated by the system (2) with Gaussian additive noise, i.e., noise distributed as (1) with $\gamma = 0$. The Gaussian maximum likelihood estimates of the parameters σ_o^2 and σ_ϵ^2 are obtained by minimizing the log likelihood function

$$l(\sigma_o^2, \sigma_\epsilon^2) = n\log(2\pi) + \sum_{i=1}^{n} \log s_i^2(\sigma_o^2, \sigma_\epsilon^2) + \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i^{i-1}(\sigma_o^2, \sigma_\epsilon^2))^2}{s_i^2(\sigma_o^2, \sigma_\epsilon^2)} \qquad (20)$$

with respect to σ_o^2 and σ_ϵ^2, where $\hat{y}_i^{i-1}$ and s_i^2 are obtained by the usual Kalman filter recursions (obtained from the robust filter with $c = \infty$).

However, the initial value P_0 is undefined for the case of pth order difference model because of its nonstationarity. To avoid this difficulty, Harvey (1981) suggests that one can choose

$$P_0 = \lambda I$$

where λ is large scalar and I is the identity matrix. Using this initial value, one can ignore the first k observations in calculating the log likelihood function for some small integer k. Throughout the paper, λ is chosen to be the square of the range of the data and $x_0 = 0$. The log likelihood function (19) is then defined as

$$l(\sigma_o^2, \sigma_\epsilon^2) = n \log(2\pi) + \sum_{i=k+1}^{n} \log s_i^2(\sigma_o^2, \sigma_\epsilon^2) + \sum_{i=k+1}^{n} \frac{(y_i - \hat{y}_i^{i-1}(\sigma_o^2, \sigma_\epsilon^2))^2}{s_i^2(\sigma_o^2, \sigma_\epsilon^2)}. \qquad (21)$$

To robustify the ML approach in the case of additive outliers which occur when $\gamma > 0$ in (1), we propose that the parameters σ_o^2 and σ_ϵ^2 be estimated by minimizing the 'robustified' log likelihood type loss function

$$\tilde{l} = \sum_{i=k+1}^{n} \log s_i^2(\sigma_o^2, \sigma_\epsilon^2) + \sum_{i=k+1}^{n} \rho_{c_1}(\frac{y_i - \hat{y}_i^{i-1}(\sigma_o^2, \sigma_\epsilon^2)}{s_i}) \qquad (22)$$

where

$$\rho_{c_1}(x) = \begin{cases} x^2 & |x| \leq c_1 \\ 2c_1|x| - c_1^2 & |x| > c_1 \end{cases}$$

is the Huber (1964) loss function with "tuning" constant c_1.

In our examples and applications to radar glint noise, we have chosen the integer k to be about 5 percent of the total number of observations and c_1 is chosen to be 1.2 as suggested by the study of Martin and Su (1985).

The exact calculation of estimates $\hat{\sigma}_o^2$ and $\hat{\sigma}_\epsilon^2$ which minimizes (22), in closed form is intractable, and one needs to employ some numerical methods for minimizing $\tilde{l}$. Since we only have to estimate two parameters, a simple brute force numerical method is possible. Evaluate the function $\tilde{l}$ on a grid of values of σ_o^2 and σ_ϵ^2, and choose the grid value that minimizes $\tilde{l}$. This is the method we use for our examples and applications.

11

When the model for the data is correct, the log likelihood function should behave nicely, that is, it is approximately convex and locally quadratic near the minimum so that the minimum is easy to be detected. Likewise, one expects the robust loss function $\tilde{l}$ to have similar behaviour if the model is correct. However, if the model is incorrect, the log likelihood and robust loss functions may exhibit poor behaviour, e.g., there may not be a clear minimum of the function. This behaviour is clearly exhibited in the application of the method to some artificial data sets and the radar glint noise data.

In the next subsection, we present two examples to illustrate how the method works. We use artificial data sets in these examples, one with the nominal variance in (1) $\sigma_o^2 = 0$ and another with $\sigma_o^2 \geq 0$. We also examine how the robust loss function $\tilde{l}$ behaves when incorrect (mismatched) models are used in these examples.

4.2   Examples

Example 1: $(\sigma_o^2 = 0)$

A sample of 200 data points is generated from a contaminated autoregressive model of order one with $\phi = 1$ (i.e. the first order difference model) and $x_0 = 0$. The model is

$$x_t = x_{t-1} + \epsilon_t$$
$$y_t = x_t + v_t$$

where ϵ_t is white noise standard Gaussian process and v_t is generated from (1) with $\sigma_o^2 = 0$, $\gamma = .1$ and $H \sim N(0, 25)$. Thus there is no nominal additive Gaussian noise to contend with, and $\tilde{l} = \tilde{l}(\sigma_\epsilon^2)$ depends on just one parameter, the innovation variance σ_ϵ^2.

The data are plotted in Figure 1.a. The plot shows that there are big spikes in the data and the nonstationarity of the data is rather obvious.

The loss function $\tilde{l}(\sigma_\epsilon^2)$ was evaluated for 100 values of σ_ϵ^2 equally spaced from .1 to 4. The results, displayed in Figure 1.b, show that the function is quite smooth and essentially convex. The minimum value of $\tilde{l}(\sigma_\epsilon^2)$ is 342.6, occurring at $\hat{\sigma}_\epsilon^2 = 1.30$.

To illustrate how the robust loss function $\tilde{l}$ behaves when the incorrect model is used, we repeat the above process with 2nd and 3rd order difference models used for constructing the predictor $\hat{y}_t^{i-1}$. The values of $\tilde{l}$ corresponding to the incorrect assumptions of 2nd and 3rd order difference models are plotted in Figures 1.c and 1.d respectively. Not only are the minimum values of $\tilde{l}$, based on incorrect assumptions of 2nd and 3rd order difference models considerably larger than that when the correct model is used, $\tilde{l}$ also looses its smoothness and convexity when the incorrect model is used. Greater roughness occurs when the model mismatch is greater. The extensive degree of roughness associated with model mismatch surprised us a bit. Perhaps this behaviour provides a tipoff that the incorrect model is being used.

To demonstrate how well the fixed-lag smoothers work using the estimated value of σ_ϵ^2, we display the data in Figure 2, along with lag-k smoothed values for k = 0, 2 and the corresponding state estimation residuals. Figure 2.a shows the observed data and the lag-0 smoother $\hat{x}_t^t$, while Figure 2.b shows the corresponding state estimation residuals $x_t - \hat{x}_t^t$. Note that under the assumption that $\sigma_o^2 = 0$, we have $y_t = x_t$ and $\hat{x}_t^t = y_t$ much of the time. Correspondingly the residuals $x_t - \hat{x}_t^t$ are zero much of the time. Results for the lag-2 smoother $\hat{x}_t^{t+2}$ are shown in Figures 2.c and 2.d.

Figures 2.a and 2.c show that the lag-0 and lag-2 smoothers follow the data quite well and they are successful in removing both moderate and large outliers. Overall, the state estimation residuals for the lag-2 smoother (Figure 2.d) tend to be smaller than those for the lag-0 smoother (Figure 2.b), as one would expect.

Example 2: $(\sigma_o^2 \geq 0)$

Again 200 observations are generated by the same model as in the first example except that now $\sigma_0^2 = 4$ and $H \sim N(0, 50)$. Thus there is a nominal additive Gaussian noise variance $\sigma_o^2 \geq 0$ which we need to estimate. So now $\tilde{l} = \tilde{l}(\sigma_\epsilon^2, \sigma_o^2)$ is a function of the two unknown variances $\sigma_\epsilon^2, \sigma_o^2$.

The data is displayed in Figure 3.a. The outliers are now a bit less obvious because of the $N(0, 4)$ Gaussian component of v_t, but they are none the less present. The loss function $\tilde{l}$ was evaluated for 100 values of σ_ϵ^2 equally spaced from .1 to 4 and 100 values of σ_0^2 equally spaced from .1 to 8. The contours of this function are plotted against σ_ϵ^2 and σ_0^2 in Figure 3.b. The plot shows that the function is quite smooth and

approximately convex so that the minimum is quite easily located. The minimum of 684.4 is obtained at $\sigma_0^2 = 5.0$ and $\hat{\sigma}_\epsilon^2 = 1.3$.

In Figures 3.c and 3.d respectively, we show the contours of the function $\tilde{l}$ based on the use of incorrect 2nd and 3rd order difference models to construct $\hat{y}_i^{i-1}$. The plots suggest that local convexity of $\tilde{l}$ is still present, and the roughness of $\tilde{l}$ associated with a mismatched model is not nearly as great as in the previous example where we set $\sigma_o^2 = 0$. For the case of an incorrectly used 2nd order model, the minimum of $\tilde{l}$ is 702.9, occurring at $\hat{\sigma}_o^2 = 6.6$, $\hat{\sigma}_\epsilon^2 = .035$. While the estimate $\hat{\sigma}_o^2$ is increasing, $\hat{\sigma}_\epsilon^2$ is decreasing and quite smaller than the true value $\sigma_\epsilon^2 = 1$. When using an incorrect third order model, the estimate $\hat{\sigma}_o^2 = 7.4$ is large compared with the true value $\sigma_o^2 = 4$; however, the really dramatic effect is the extremely small value $\hat{\sigma}_\epsilon^2 = .0006$. These results indicate that when a mismatched model of higher order is used, the observation variance will be overestimated and the innovation variance will be underestimated, and that this effect will be greater when the degree of mismatch is greater.

Note that the larger the observation variance is, the "smoother" the conditional expectation $\hat{y}_i^{i-1}$ would be. In addition, the use of higher order difference model intrinsically leads to more smoothness.

In Figure 4, we plot the observations along with the lag-k smooths, and the state estimation residuals for k = 0, 2. The results are as one might expect. The lag-2 smoother (Figure 4.c) shows much more smoothness than the lag-0 smoother (Figure 4.a). The state estimation residuals from the lag-2 smoother are also noticeably smaller than those for the lag-0 smoother.

5 Application to Glint Noise Data

5.1 Description of Data

In Figure 5, we display a segment of radar glint noise data encountered in radar tracking of aircraft targets. The ordinate values are the apparent line-of-sight angles of the aircraft (in degrees), measured from baresite of the fixed position radar dish. The origin of the ordinate corresponds to a right-angle side view of the aircraft situated in a level position. The exceedingly spiky outlier prone behaviour of this process is due to interference from the multiple returns of the radar signal reflecting from the various components of the aircraft structure (e.g., wing leading edges, tails, cockpit, etc.). The spikes are often huge, sometimes giving a misindicator of true angular position by as much as 100° or more. In spite of having an extremely non-Gaussian character (see Hewer, Martin and Zeh, 1987), the process has often been treated as being Gaussian in engineering analyses. Needless to say, such analyses may be quite misleading.

The radar glint noise also contains a low-frequency trend-like component which one can see in Figure 5 (eyeball a smooth curve through the data). This low-frequency component, called "bright spot wander", represents the electromagnetic center of gravity of the aircraft, which obviously changes (slowly) as the aircraft rotates.

5.2 Model Estimation

Prior experiments (Martin and Thompson, 1982) with robust autoregressive model fitting and robust smoothing indicate that the trend-like component in radar glint noise is well modeled by 1st, 2nd or at most 3rd order autoregressions with parameters on the boundary of nonstationarity. Thus, for this data it is natural to use the simple approach based on 1st, 2nd and 3rd order difference models and robust estimation of the parameters σ_ϵ^2 and σ_0^2. We examine two cases, one with the assumption $\sigma_0^2 = 0$ and another without this assumption. Employing the estimation procedure described in Section 4, we obtained the following results.

Assuming that $\sigma_0^2 = 0$, the robust loss function l for 1st, 2nd and 3rd order

difference models, are evaluated and plotted in Figure 6, (a)-(c) for 1st through 3rd order difference models, respectively. The estimates $\hat{\sigma}_\epsilon^2$ and the values of the loss function $\tilde{l}(\hat{\sigma}_\epsilon^2)$ are presented in Table 1.

Table 1: The estimates $\hat{\sigma}_\epsilon^2$ and $\tilde{l}(\hat{\sigma}_\epsilon^2)$.

Order	$\hat{\sigma}_\epsilon^2$	$\tilde{l}(\hat{\sigma}_\epsilon^2)$
1st	177	3997.4
2nd	205	4468.8
3rd	389	5206.3

On the basis of minimum $\tilde{l}$, one is clearly led to prefer the first order model. We note also that whereas the function $\tilde{l} = \tilde{l}(\sigma_\epsilon^2)$ is rather smooth and nearly convex for the 1st order model, $\tilde{l}$ becomes increasingly rough and non-convex for the 2nd and 3rd order models. This is further evidence that the 2nd and 3rd order models are inappropriate.

The glint noise data and the lag-3 smooths corresponding to the estimated 1st, 2nd and 3rd order difference models are plotted in Figure 7 (a rationale for the choice of lag-3 is provided in the next section). We plot only the first 300 data points so that the behaviour of the smoothers can be easily examined. Figure 7.a shows that for the first order model, the lag-3 smoother is quite successful in chopping off the large spikes/outliers, and otherwise follows the data. This is as expected because of the assumption $\sigma_0^2 = 0$, according to which a "data-cleaner" behaviour is expected (see, Martin and Thompson, 1982). However, the smoother perhaps does not chop off the more moderate sized spikes as effectively as one might wish.

Figure 7.b shows that for the second order model, the lag-3 smoother does not behave quite so well. It overshoots in a few places, and is consequently not as effective in chopping off the spikes. The behaviour of the lag-3 smoother for the 3rd order model, shown in Figure 7.c, is even worse in this regard. The behaviour of the lag-3 smoothers for the 2nd and 3rd order difference models are consistent with our observation that these models fit poorly using the minimized $\tilde{l}$ as a criterion, under

16

the assumption that $\sigma_o^2 = 0$.

We now consider the case where $\sigma_o^2 \geq 0$ is treated as an unknown, and the two parameters σ_0^2 and σ_ϵ^2 are estimated by minimizing the robust loss function $\tilde{l}$. The estimates $\hat{\sigma}_0^2$ and $\hat{\sigma}_\epsilon^2$ and the values of the loss function $\tilde{l}(\hat{\sigma}_0^2, \hat{\sigma}_\epsilon^2)$ are presented in Table 2.

Table 2: The estimates $\hat{\sigma}_o^2$, $\hat{\sigma}_\epsilon^2$, and $\tilde{l}(\hat{\sigma}_o^2, \hat{\sigma}_\epsilon^2)$.

Order	$\hat{\sigma}_o^2$	$\hat{\sigma}_\epsilon^2$	$\tilde{l}(\hat{\sigma}_o^2, \hat{\sigma}_\epsilon^2)$
1st	57.2	73.10	3968.5
2nd	129.4	1.62	4073.3
3rd	154.1	.04	4176.6

The contours of $\tilde{l}$ are plotted in Figures 8a, b, c for the 1st, 2nd and 3rd order difference models respectively. As in the artificial Example 2 of Section 4, the surface appears most smooth and nearly convex for the first order model, with some degradation in this respect for the 2nd and 3rd order models. On the basis of minimum $\tilde{l}$, the first order model is again preferred. Using this model, the robust loss function attains a minimum value of 3968.5, which is smaller than the minimum value of 3997.4 obtained in the case when it was assumed that $\sigma_o^2 = 0$. Thus from the robust loss function point of view the assumption $\sigma_o^2 \geq 0$ leads to a better model fit.

The original glint noise data (first 300 points), and results of applying the lag-3 smooths are displayed in Figures 9a-c for the 1st-3rd order models, respectively. The lag-3 smooths in Figure 9a for the first order difference with $\hat{\sigma}_o^2 = 57.2$ exhibit more smoothness than those in Figure 7a for a first model difference model with the assumption $\sigma_o^2 = 0$. Note however that for the 2nd and 3rd order models, the combination of higher order and larger ratio $\hat{\sigma}_o^2/\hat{\sigma}_\epsilon^2$ of estimated variances results in considerably increased smoothness.

Actually, if one takes a "smoothing for visual display" point of view, the second order model lag-k smoother shown in Figure 9.b provides a most pleasing presentation: It fits a more smooth curve through the data than in the case of the first order differ-

ence model, and the use of a third order difference model yields little improvement over the second order model.

6 Choice of Lag

In the previous section we displayed results for lag-3 smoothing of radar glint noise. The choice of lag 3 for this particular example was determined empirically by examining the quality of the smoother for various lags. The reason for this is that an analytical approach for choosing k is not feasible for robust smoothers because of their nonlinear structures (by way of contrast, in the case of the linear Kalman smoother, the lag k might be chosen to achieve a mean-squared error within a specified factor of the fixed interval smoother mean squared error, the latter being the best that one can do).

Figures 10 and 11 display robust lag-k smoothers for k = 0, 1, 3, 5 for the first and second order difference models respectively where both σ_ϵ^2 and σ_o^2 are estimated. For the first order model there is not a great deal of difference in the quality of the smooth for various lags, there being only a slight amount of increased smoothness with increasing lag. For the second order model there is a noticeable increase in smoothness with increasing lag k, and most of this increased smoothness seems to be achieved at k = 3. Thus we displayed the k = 3 results for both first and second order models in the previous section.

To study the behaviour of the fixed-lag smoothers for different lags in a more systematic way, we carry out a simulation study as follows. We generated 50 time series of length 200 from the process

$$y_t = x_t + v_t$$

where x_t is a first order autoregression with standard normal innovations ϵ_t, and v_t is a contamination process containing outliers.

In this simulation experiment we introduce a new consideration, namely the possibility of patchy outliers of various lengths. The basic question to be addressed in this regard is whether the use of longer lags k in the lag-k robust smoother would be

helpful in dealing with patchy outliers. Thus we construct outlier patches of lengths 1, 2 and 3 by the following way. For each sample path, the process x_t are contaminated by non-overlapped patches of outliers, independently generated from $N(0, 25)$. The positions of the patches are randomly chosen. For all three patch lengths, there are 20% of outliers in total.

The robust lag-k smoothers performance is evaluated by computing for each sample path the mean-squared error for smoothing at outlier positions (MSE_1) and non-outlier positions (MSE_2) separately:

$$MSE_1 = \frac{1}{\sum_{t:v_t \neq 0} 1} \sum_{t:v_t \neq 0} (\hat{x}_t^{t+k} - x_t)^2$$

$$MSE_2 = \frac{1}{\sum_{t:v_t = 0} 1} \sum_{t:v_t = 0} (\hat{x}_t^{t+k} - x_t)^2.$$

The above mean-squared errors are computed for each of the 50 sample paths.

The boxplots of both MSEs for lag k, k = 0, 1,...,5 are displayed in Figure 12 for patches of length 1, 2 and 3. The results for MSE_1 are shown in Figure 12a, b, c, while the MSE_2 results are shown in Figure 12e, f, g.

These plots show that:

(a) Both MSE_1 and MSE_2 increase with increasing patch length.

(b) MSE_1 has a nearly symmetric distribution (with an upper tail slightly heavier than the lower tail) for patches of length one, and an increasingly asymmetric distribution skewed to the right for increasing patch length.

(c) MSE_2 has an asymmetric distribution, with an upper tail considerably heavier than the lower tail, for all 3 patch lengths.

(d) For all these patch lengths both MSE_1 and MSE_2 essentially reach their optimum, namely a distribution must compact toward zero, by lag k = 3.

(e) MSE_2 does not settle to zero with increasing lag.

The behaviour (e) has to do with the fact that the presence of outliers has an effect on the smoother at non-outliers positions. While this behaviour is not entirely

19

unavoidable, one wonders whether robust lag-k smoothers can be constructed with smaller mean-squared error at non-outlier positions.

The result (d) is consistent with our choice of k = 3 for the 1st order difference model lag-k smoothing of radar glint noise in Section 5. To make sure that k = 3 is reasonable for the 2nd order difference model we should probably run the above experiment based on such a model. More generally, this type of simulation would appear to provide useful guidelines for selecting k for robust lag-k smoothing for a wide range of state space models.

7 Sensitivity To Model Mismatch

We have proposed a simple approach to robust smoothing which involves estimation of only two parameters by virtue of assuming that the core part of the data is well modeled by low order difference models, thereby avoiding the estimation of autoregression (AR) or autoregression-moving average (ARMA) parameters. One wonders how the method will work when the core part of the data in fact comes from a stationary AR or ARMA model, with parameters "reasonably near" the boundary of nonstationary. To get a feeling for how our approach would work under such "mismatch" conditions with regard to the estimation of σ_ϵ^2 and σ_o^2, we carry out a simulation study for the AR(1) case.

Specifically we generate x_t as a first-order Gaussian autoregression with $\phi = .5$, .9, .95, .98, 1 and standard normal innovations ϵ_t, and use the estimation approach of Section 4 assuming a first difference model. The x_t are contaminated with additive outliers v_t which has the distribution (1) with nominal variance $\sigma_0^2 = 0$ (and so we do not estimate σ_o^2) and $H \sim N(0, 25)$. The contamination fraction γ is chosen to be 0, .1, .2 and .3. For each combination of γ and ϕ, 100 time series of length 200 are generated. For each series, the robust loss function $\tilde{l}(\sigma_\epsilon^2)$ is evaluated, and the estimate $\hat{\sigma}_\epsilon^2$ which minimizes the loss function is obtained. The means and standard deviations of the estimates $\hat{\sigma}_\epsilon^2$ are presented in Table 3.

The results show that the means (and of course the standard deviations as well) increase with an increasing fraction contamination γ. The means and the standard

Table 3: Estimates of $E\hat{\sigma}_\epsilon^2$ obtained via Monte Carlo (sample standard deviations of estimates in paratheses)

True ϕ	$\gamma = 0$	$\gamma = .1$	$\gamma = .2$	$\gamma = .3$
0.50	.91 (.16)	1.51 (.36)	2.24 (.48)	3.41 (.96)
0.90	.92 (.13)	1.43 (.27)	2.20 (.47)	3.45 (.82)
0.95	.94 (.12)	1.47 (.27)	2.28 (.53)	3.40 (.82)
0.98	.96 (.14)	1.47 (.25)	2.27 (.57)	3.39 (.88)
1.00	.93 (.13)	1.46 (.29)	2.22 (.50)	3.33 (.77)

deviations are relatively constant with respect to changing ϕ. This suggests that our method is quite insensitive to model mismatch as far as estimation of σ_ϵ^2 goes, when it is known that $\sigma_o^2 = 0$. Except for the case 0% contamination in which the estimated values are biased downward, the estimated values $\hat{\sigma}_\epsilon^2$ for other cases are biased upward. Although the biases in the estimates of σ_ϵ^2 appear to be large when $\gamma > 0$, the biases of $\hat{\sigma}_\epsilon$ as an estimate of the standard deviation σ_ϵ are not nearly so large.

It is customary in robust estimation that biased estimates are standardized so as to be unbiased at the nominal model (i.e., $\phi = 1$ and $\gamma = 0$ in this case). This standardization is accomplished in the present case by dividing the estimate $\hat{\sigma}_\epsilon^2$ by its asymptotic value $E_\infty \hat{\sigma}_\epsilon^2$. To do this one needs to know the asymptotic value $E_\infty \hat{\sigma}_\epsilon^2$ of the estimate at the nominal model. Sometimes, it is possible to calculate such values analytically or using numerical integration (see, e.g., Hampel et al, 1986). This is not possible in the present case. However, asymptotically a sufficiently accurate estimate could be obtained by a very long simulation study in this case.

One should also study the effect of model mismatch on (i) joint estimation of σ_ϵ^2 and σ_o^2, and (ii) lag-k smoothing mean-squared error. We expect to do this in the near future.

8 Concluding Remarks

For on-line applications of our method, the crude grid-search approach to optimization needs to be replaced by an on-line gradient or Newton type approach. This may be accomplished via a stochastic approximation type algorithm (see for example, Albert and Gardner, 1967).

In view of the perhaps uncomfortably large biases for $\hat{\sigma}_\epsilon^2$ exhibited in Table 3 for $\gamma > 0$, we wonder whether a more robust estimate of σ_ϵ^2 can be obtained by some modification of the robust loss function approach described in Section 4.

References

[1] Albert, A.E. and Gardner, L.A. (1967) *Stochastic Approximation and Nonlinear Regression*. The M.I.T. Press, Cambridge, Massachusetts.

[2] Anderson, B.D.O. and Moore, J.B. (1979) *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, NJ.

[3] Fraiman, R. and Martin, R.D. (1988) "Approximate conditional mean filter for non-Gaussian noise", Technical Report, University of Washington.

[4] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986) *Robust Statistics: The Approach Based on Influence Functions*. Wiley.

[5] Harvey, A.C. (1981) *Time Series Models*. Halsted Press, New York.

[6] Hewer, G.A., Martin, R.D. and Zeh, J.E. (1987) "Robust preprocessing of Kalman filtering of glint noise", IEEE.-Trans. on Aerospace and Electronic System, AES-23, No. 1, pp. 120-128.

[7] Huber, P.J. (1964) "Robust estimation of a location parameter", Annals of Mathematical Statistics. 35, pp. 73-101.

[8] Huber, P.J. (1981) *Robust Statistics*. Wiley, New York, NY.

[9] Jazwinski, A.H. (1970) *Stochastic Processes and Filtering Theory.* New York: Academic Press.

[10] Martin, R.D. (1979) "Approximate conditional-mean type smoothers and interpolators", in *Smoothing Techniques for Curve Estimation,* edited by T. Gasser and M. Rosenblatt. Springer, Berlin, pp. 117-143.

[11] Martin, R.D. (1981) "Robust smoothers and outliers interpolators". Unpublished manuscript.

[12] Martin, R.D. and Su, K.Y. (1985) "Robust Filters and Smoothers: Definition and Design", Technical Report No. 58, University of Washington.

[13] Martin, R.D. and Thomson, D.J. (1982) "Robust-Resistant Spectrum Estimation", Proceedings of the IEEE, Vol. 70, No. 9, pp. 1097-1115.

[14] Masreliez, C.J. (1975) "Approximate non-Gaussian Filtering with Linear State and Observation Relation", IEEE-Trans. Automat. Contr., Vol. AC-20, pp. 361-371.

[15] Meditch, J.S. (1969) *Stochastic Optimal Linear Estimation and Control.* New York: McGraw-Hill.

[16] Moore, J.B. (1973) "Discrete-Time Fixed-Lag Smoothing Algorithms", Automatica, Vol. 9, No. 2, pp. 163-173.

[17] Stuck, B.W. (1976) "Minimum Error Dispersion Linear Filtering of Symmetric Stable Processes", IEEE-Trans. Automat. Contr., Vol. 17, pp. 507-509.

[18] Wecker, W.E. and Ansley, C.F. (1983) "The Signal Extraction Approach to Nonlinear Regression and Spline Smoothing", JASA, 78, pp. 81-89.

FIGURE 1: DATA AND ROBUST LOSS FUNCTION
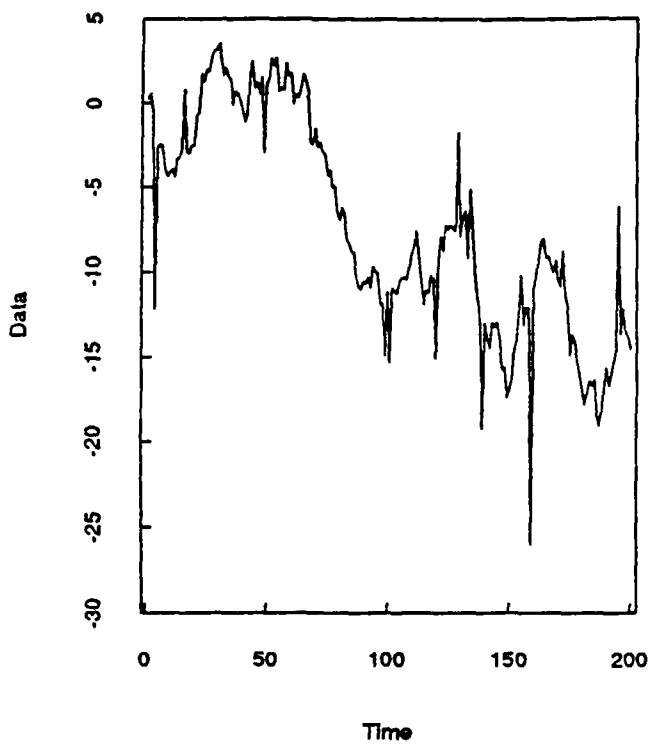
(a) GAUSSIAN AR(1) SERIES WITH
ADDITIVE OUTLIERS
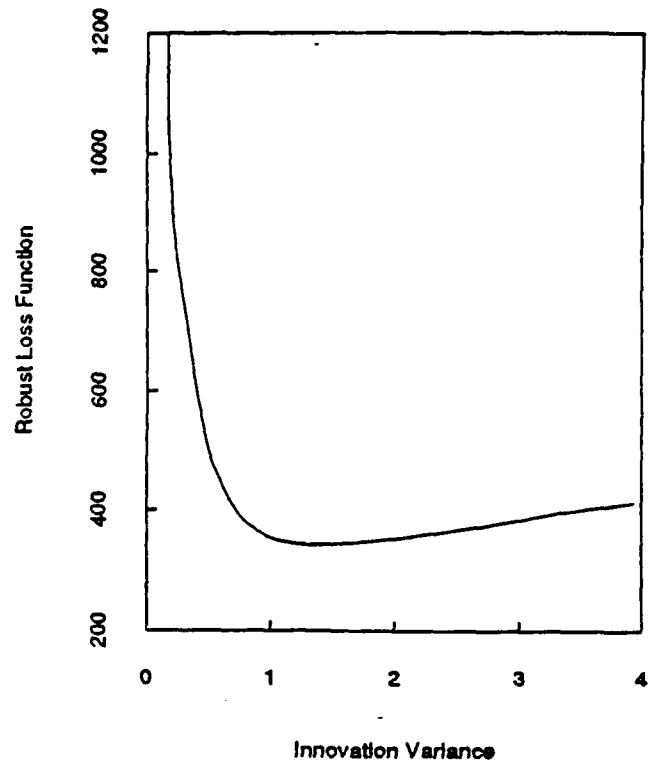
(b) FIRST ORDER DIFFERENCE MODEL
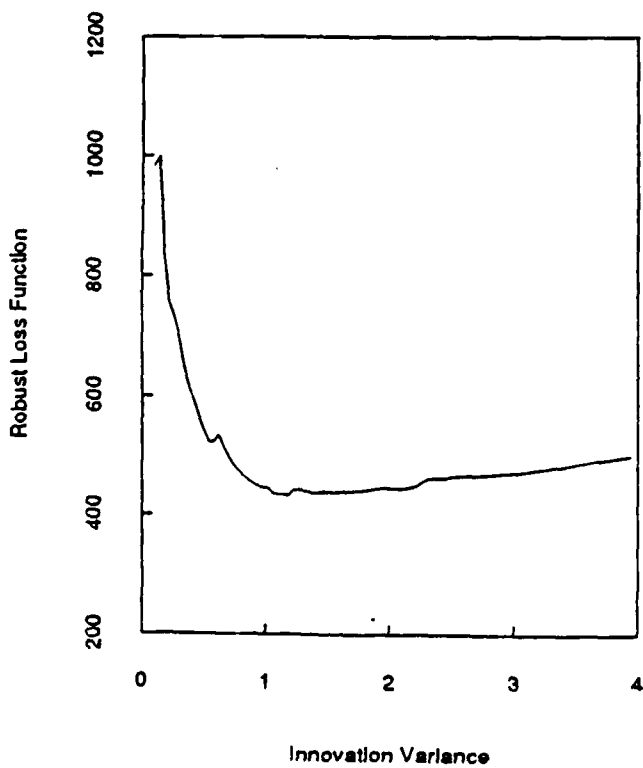
(c) SECOND ORDER DIFFERENCE MODEL
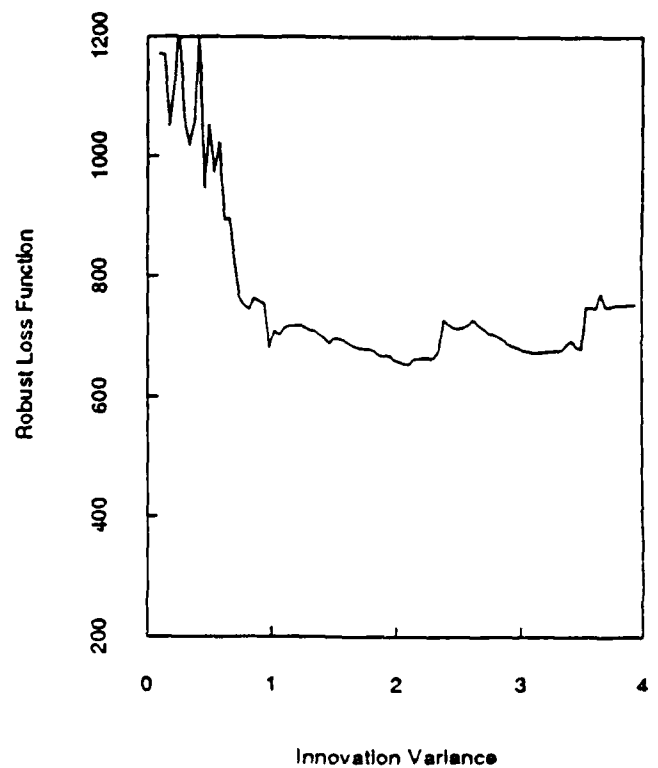
(d) THIRD ORDER DIFFERENCE MODEL

Figure 2: DATA (---) AND FIXED-LAG SMOOTHERS (—)

(a) LAG 0 SMOOTHER

(b) STATE ESTIMATION RESIDUALS FOR LAG 0 SMOOTHER

(c) LAG 2 SMOOTHER

(d) STATE ESTIMATION RESIDUALS FOR LAG 2 SMOOTHER
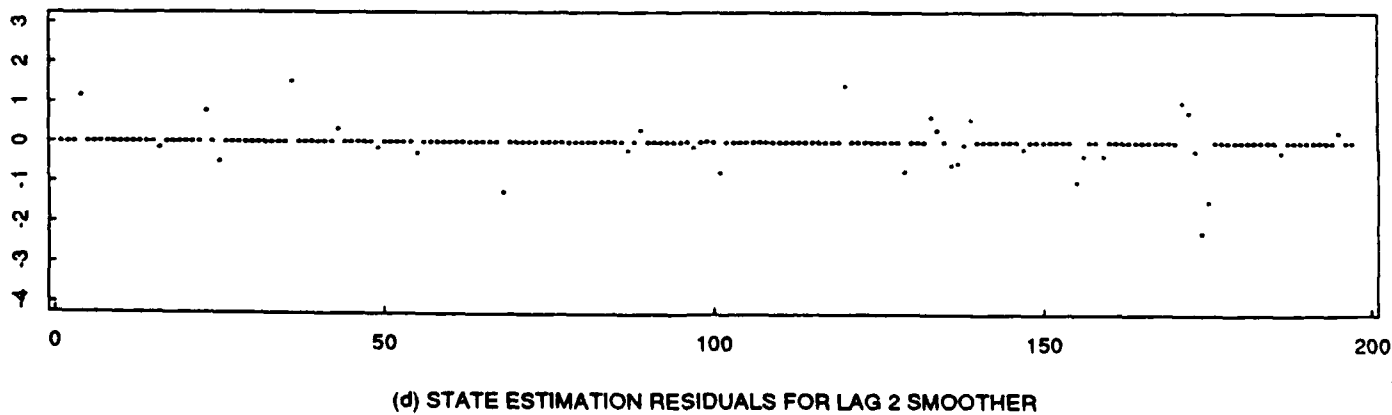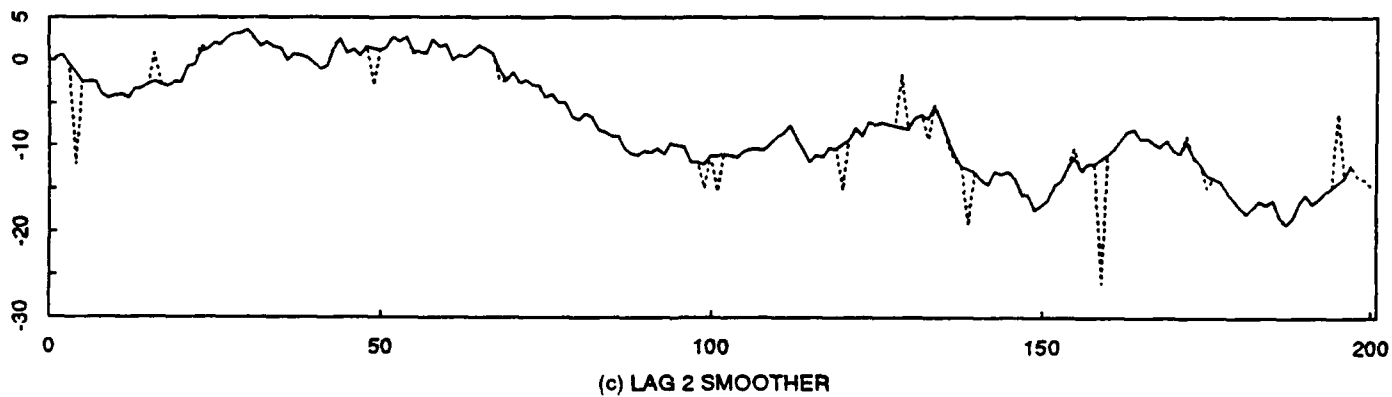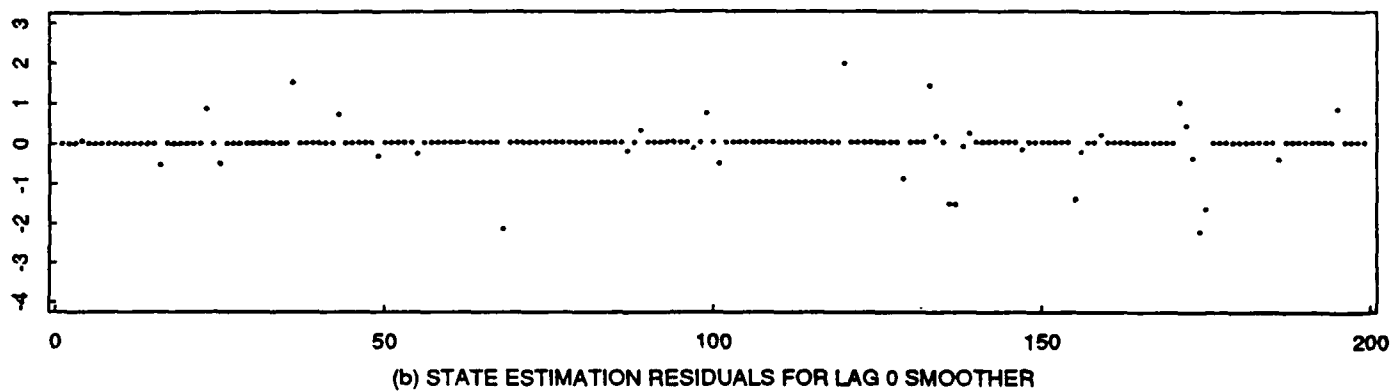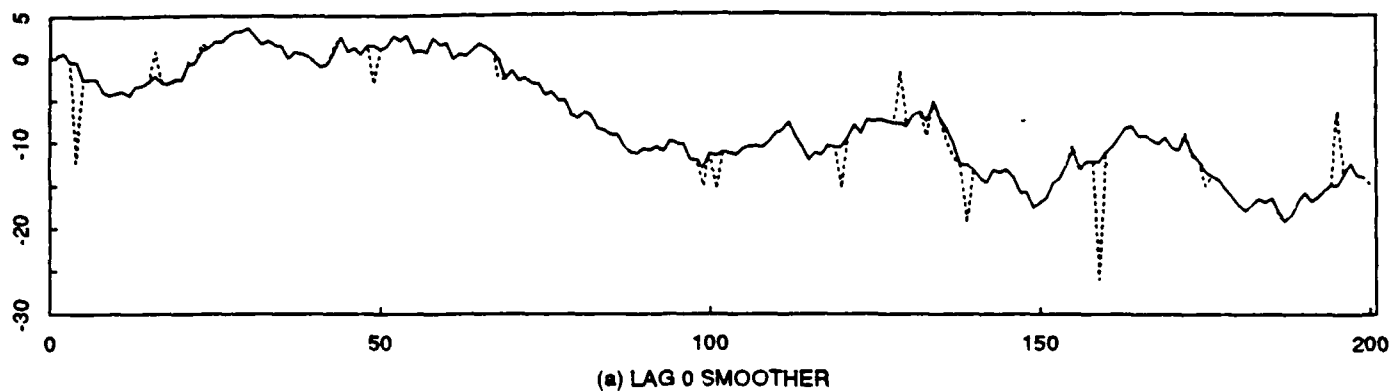
FIGURE 3: DATA AND CONTOURS OF THE ROBUST LOSS FUNCTIONS
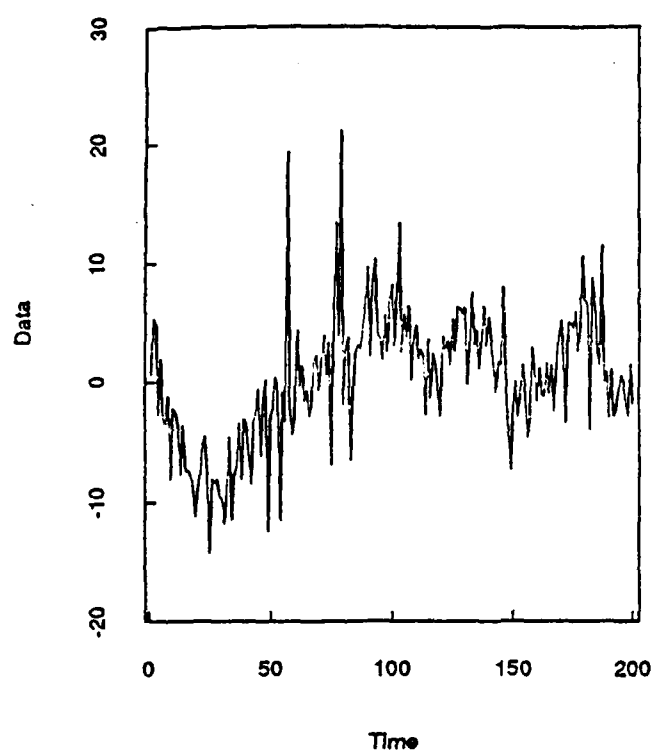
(a) GAUSSIAN AR(1) SERIES WITH ADDITIVE
PLUS CONTAMINATED GAUSSIAN NOISE
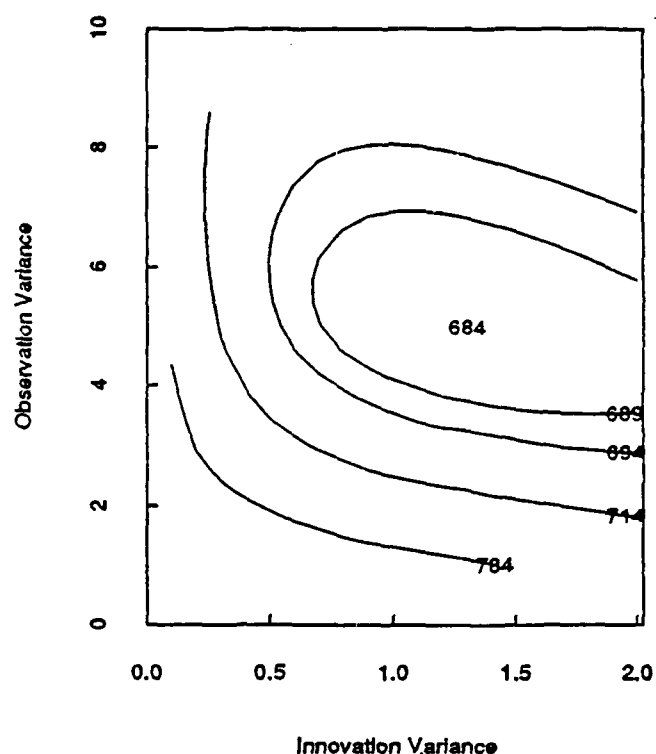
(b) FIRST ORDER DIFFERENCE MODEL
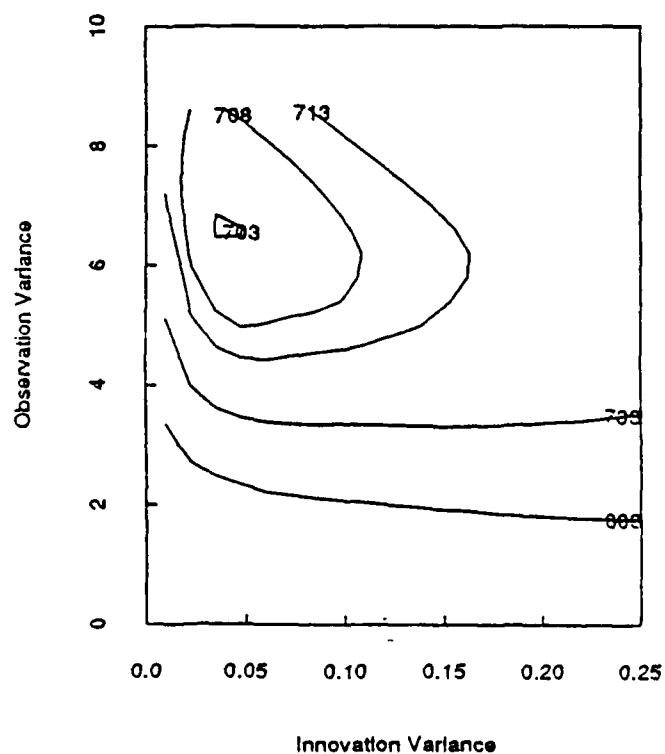
(c) SECOND ORDER DIFFERENCE MODEL
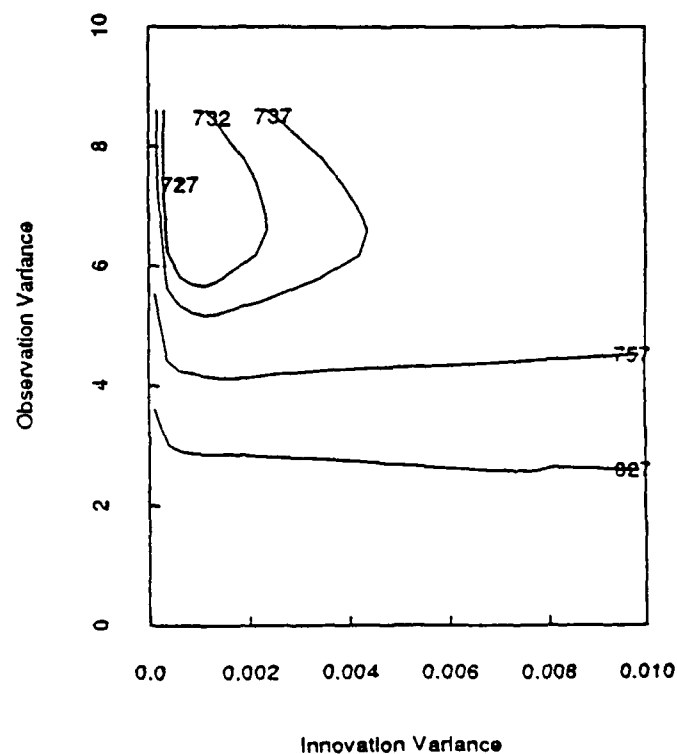
(d) THIRD ORDER DIFFERENCE MODEL

Figure 4: DATA (---) AND FIXED-LAG SMOOTHERS (—)

(a) LAG 0 SMOOTHER

(b) STATE ESTIMATION RESIDUALS FOR LAG 0 SMOOTHER

(c) LAG 2 SMOOTHER

(d) STATE ESTIMATION RESIDUALS FOR LAG 2 SMOOTHER
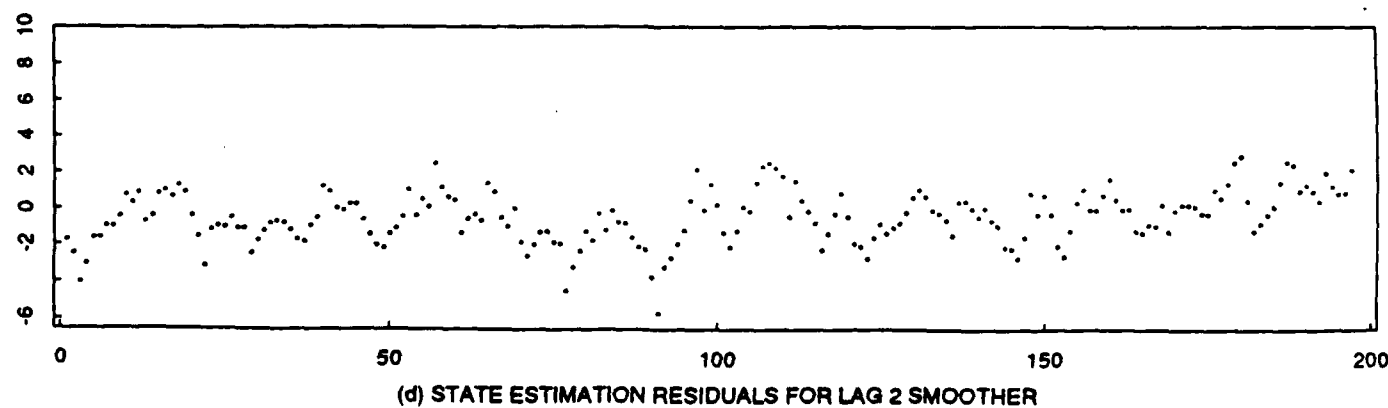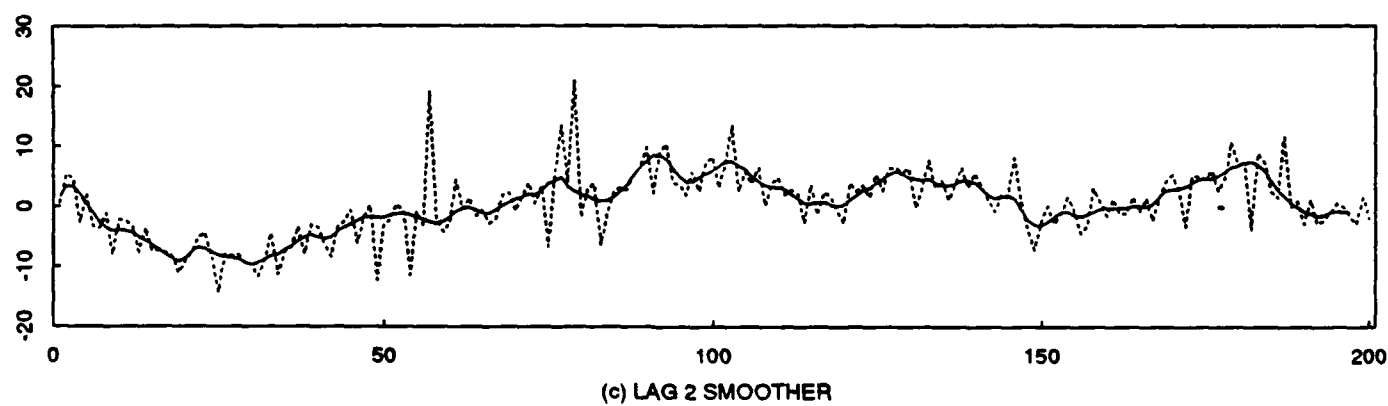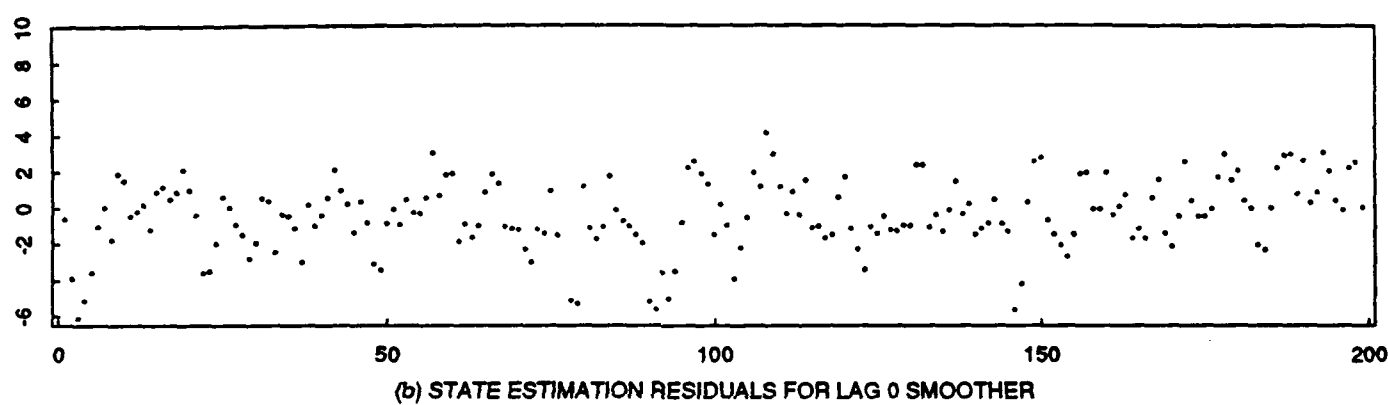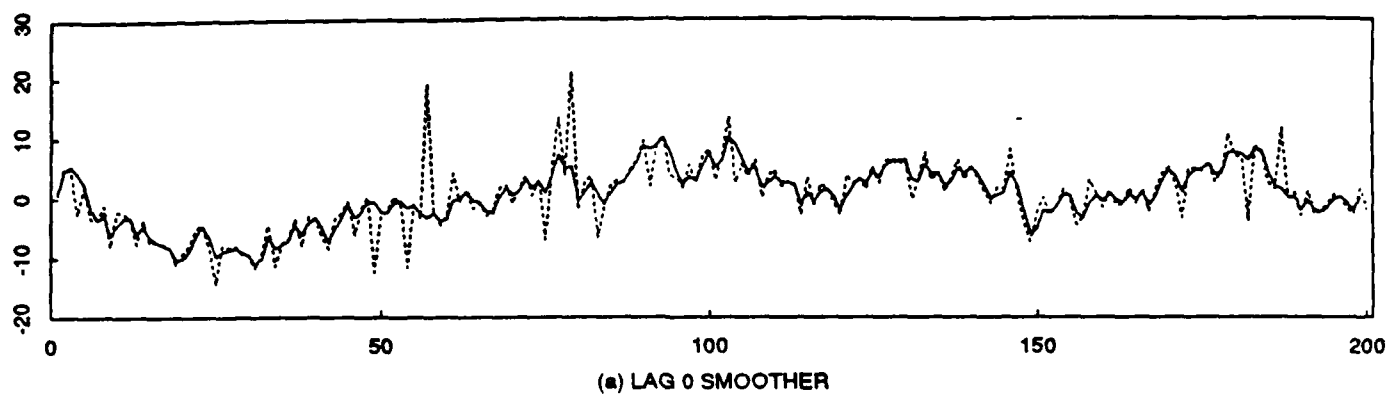
Figure 5: RADAR GLINT NOISE

Time
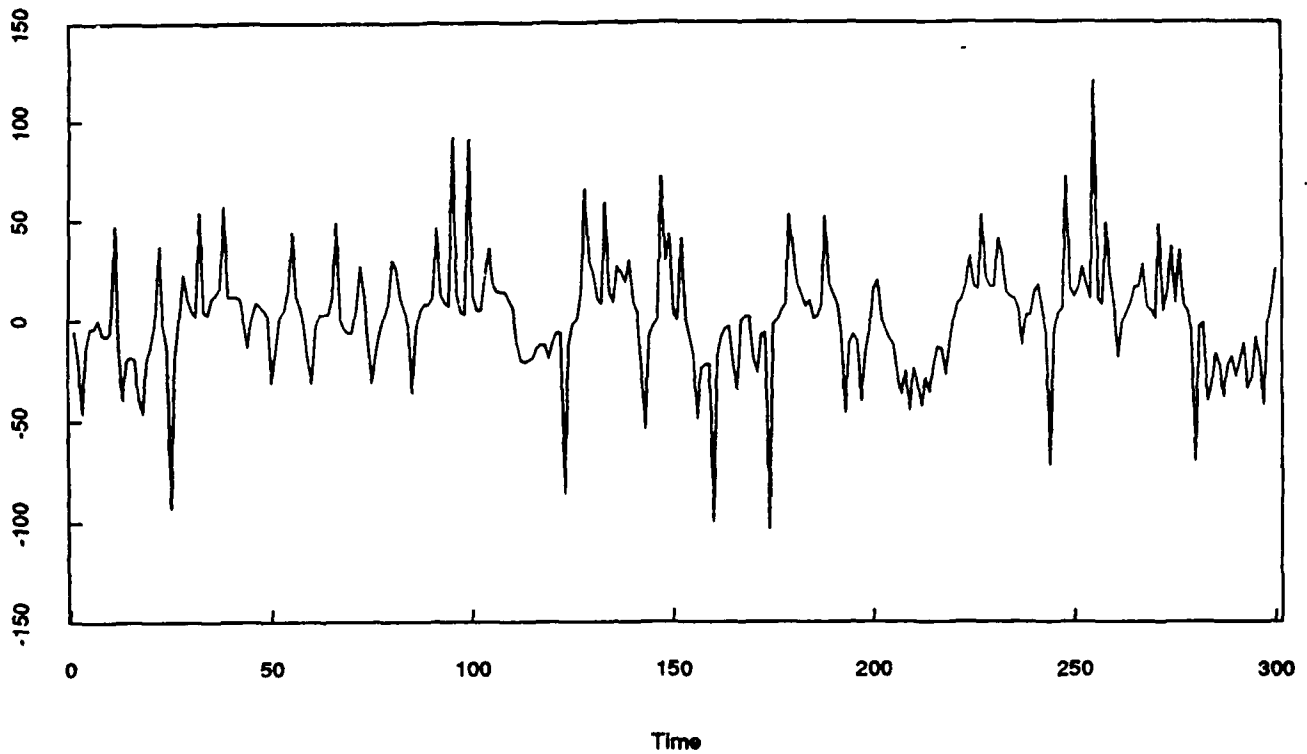
(a) THE FIRST 300 DATA POINTS

Time
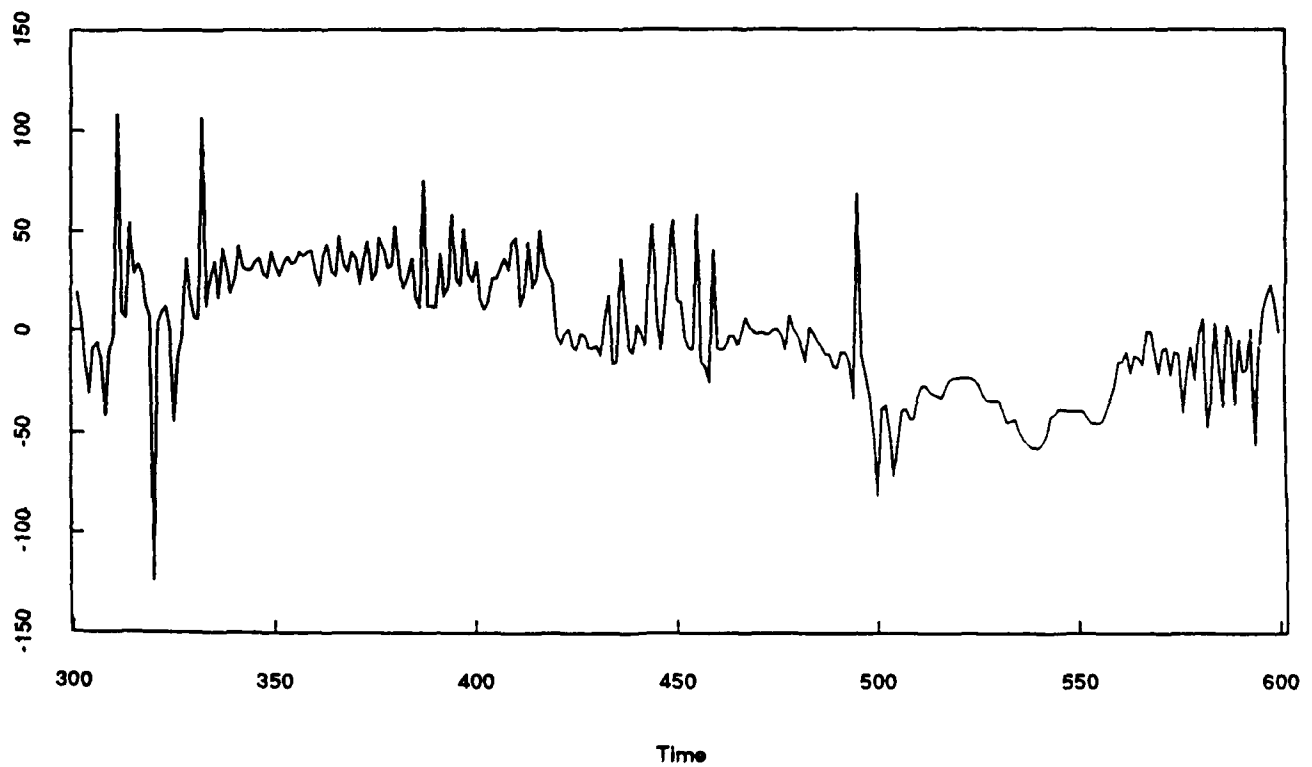
(b) THE SECOND 300 DATA POINTS

FIGURE 6: THE ROBUST LOSS FUNCTION FOR RADAR GLINT NOISE
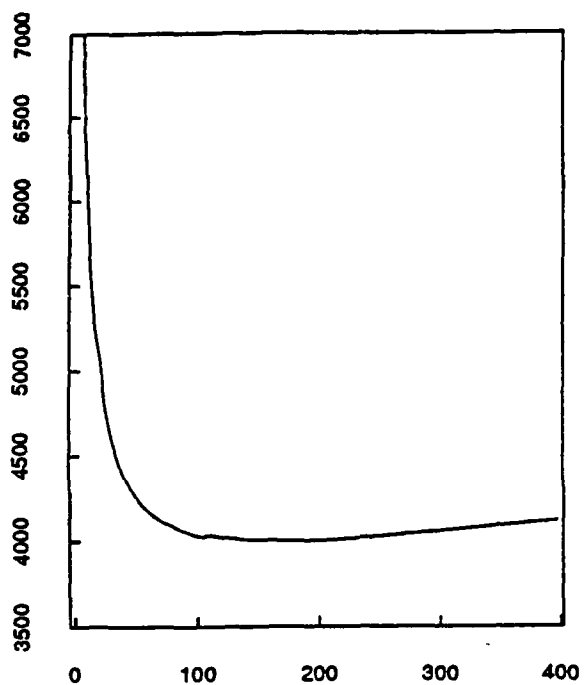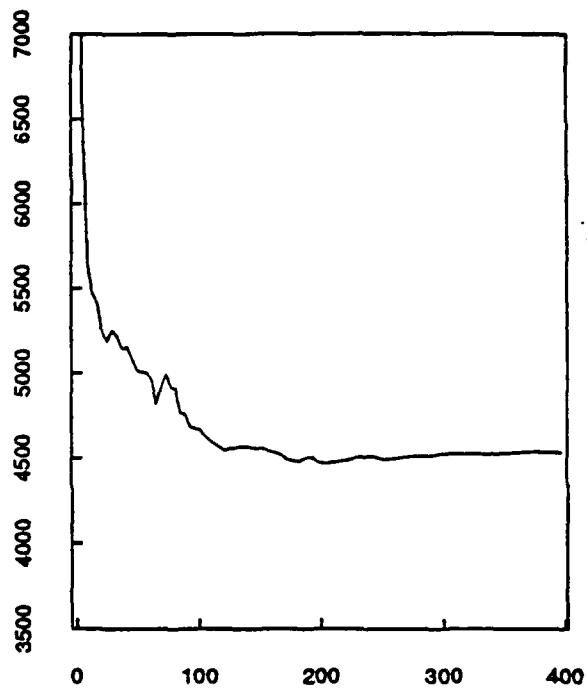
(a) FIRST ORDER DIFFERENCE MODEL
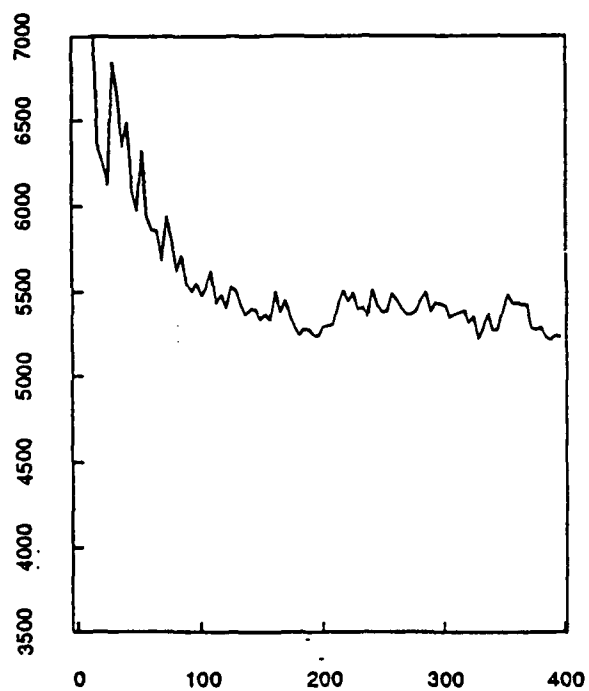
(b) SECOND ORDER DIFFERENCE MODEL

(c) THIRD ORDER DIFFERENCE MODEL

Figure 7: RADAR GLINT NOISE (---) AND LAG-3 SMOOTHER (—)

(a) FIRST ORDER DIFFERENCE MODEL, ESTIMATED INNOVATION VARIANCE = 177
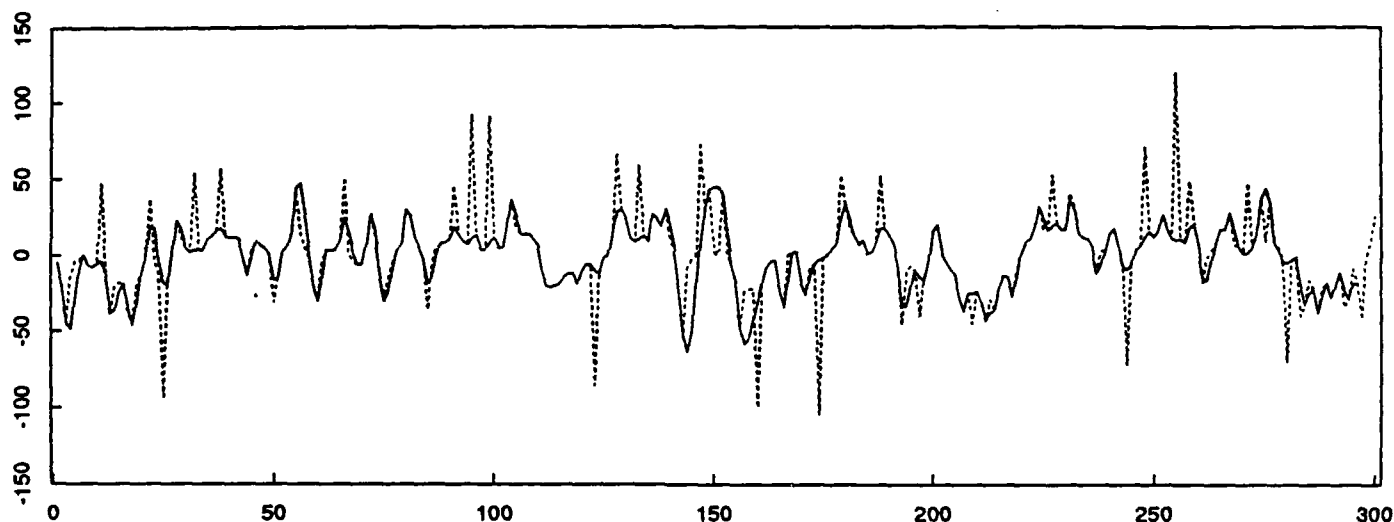
(b) SECOND ORDER DIFFERENCE MODEL, ESTIMATED INNOVATION VARIANCE = 205
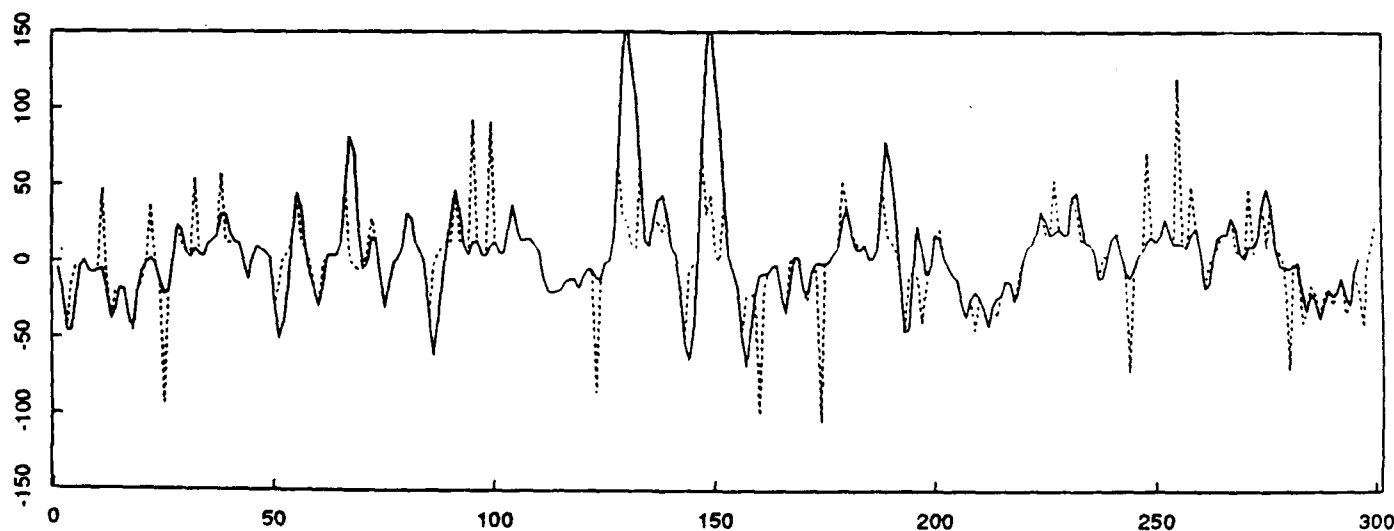
(c) THIRD ORDER DIFFERENCE MODEL, ESTIMATED INNOVATION VARIANCE = 389

FIGURE 8: CONTOURS OF THE ROBUST LOSS FUNCTION FOR RADAR GLINT NOISE
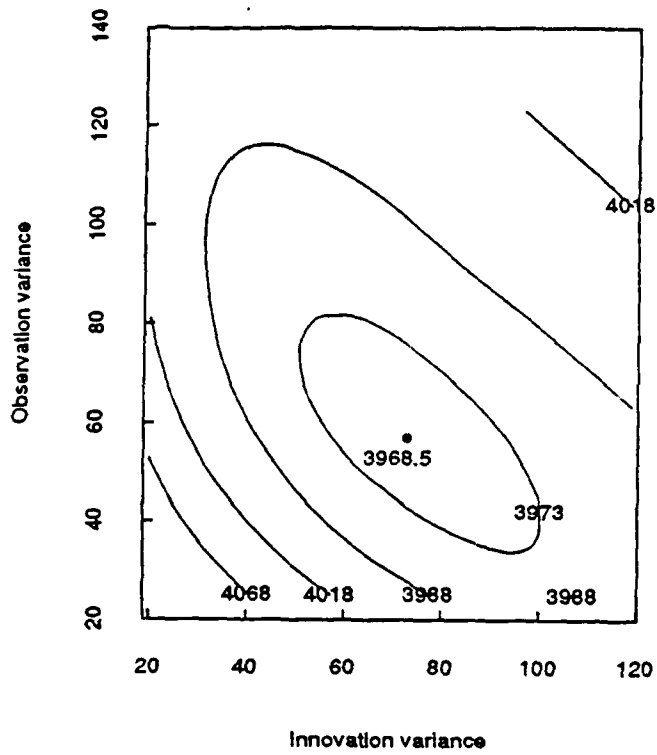
(a) FIRST ORDER DIFFERENCE MODEL
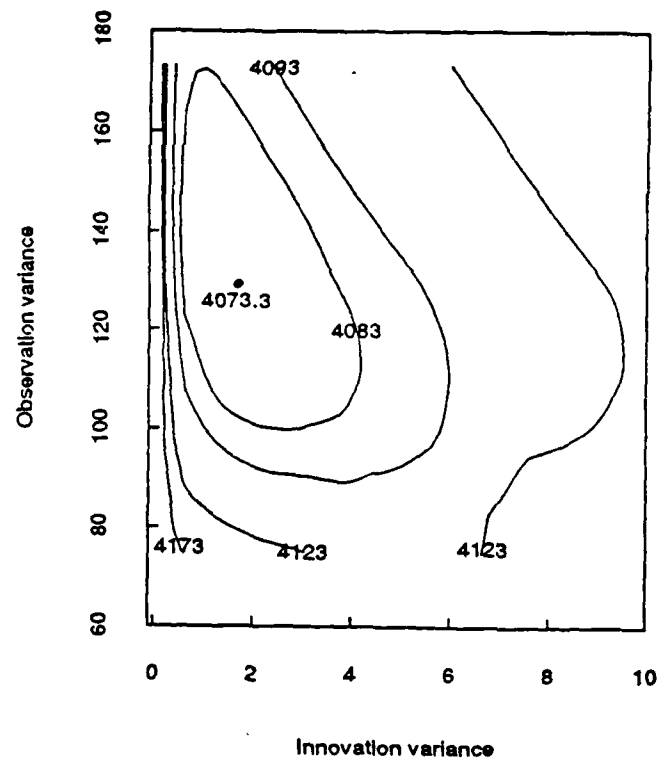
(b) SECOND ORDER DIFFERENCE MODEL
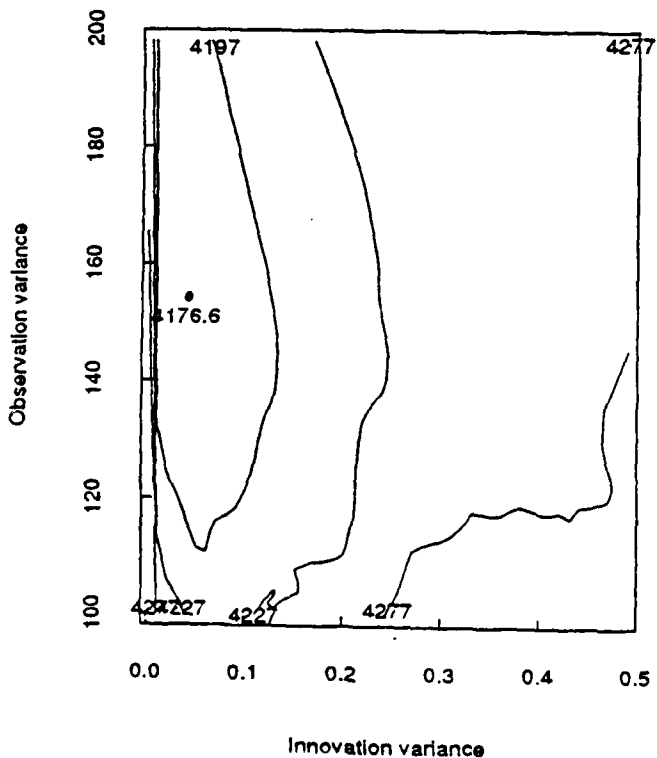
(c) THIRD ORDER DIFFERENCE MODEL

Figure 9: RADAR GLINT NOISE (---) and LAG-3 SMOOTHER (—)
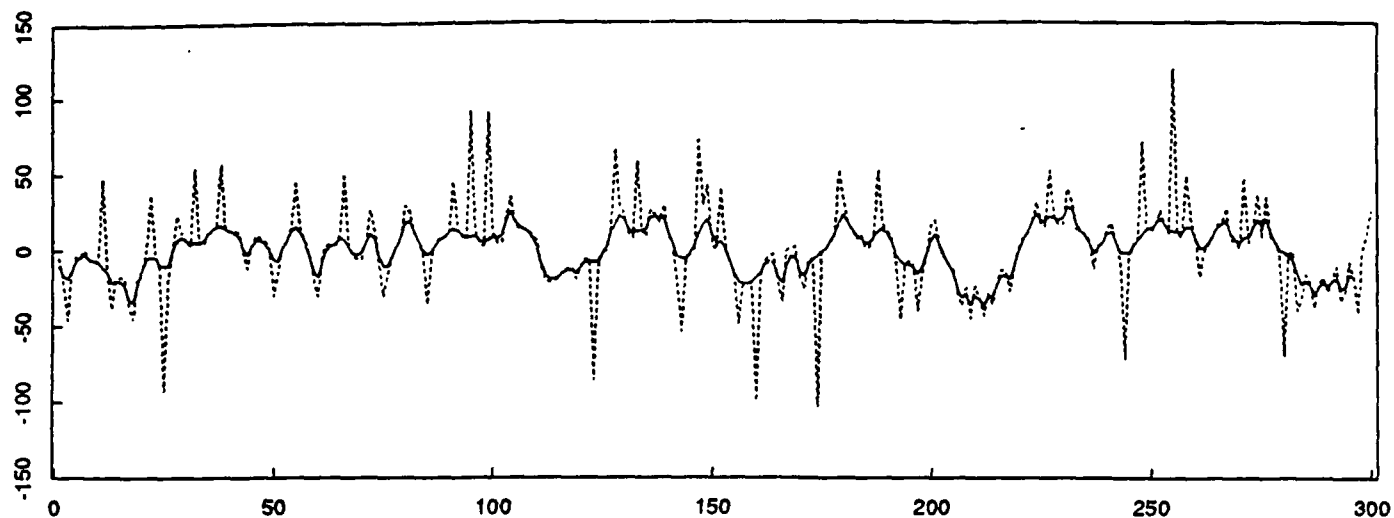
(a) FIRST ORDER DIFFERENCE MODEL, ESTIMATED INNOVATION VARIANCE = 73.1, OBSERVATION VARIANCE = 57.2
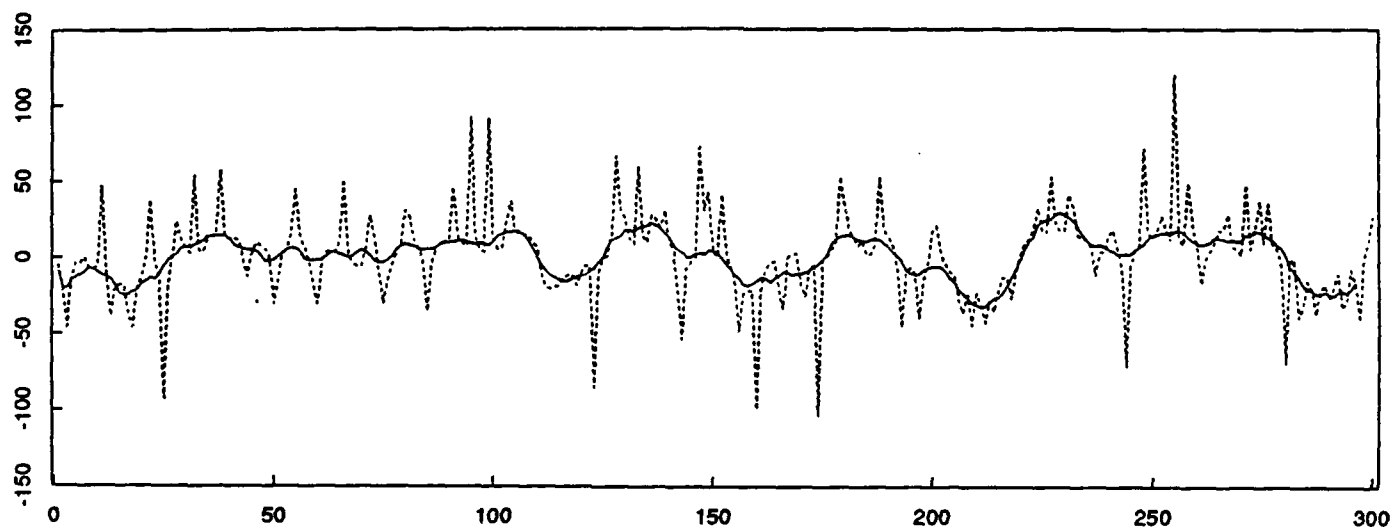
(b) SECOND ORDER DIFFERENCE MODEL, ESTIMATED INNOVATION VARIANCE = 1.62, OBSERVATION VARIANCE = 129.4
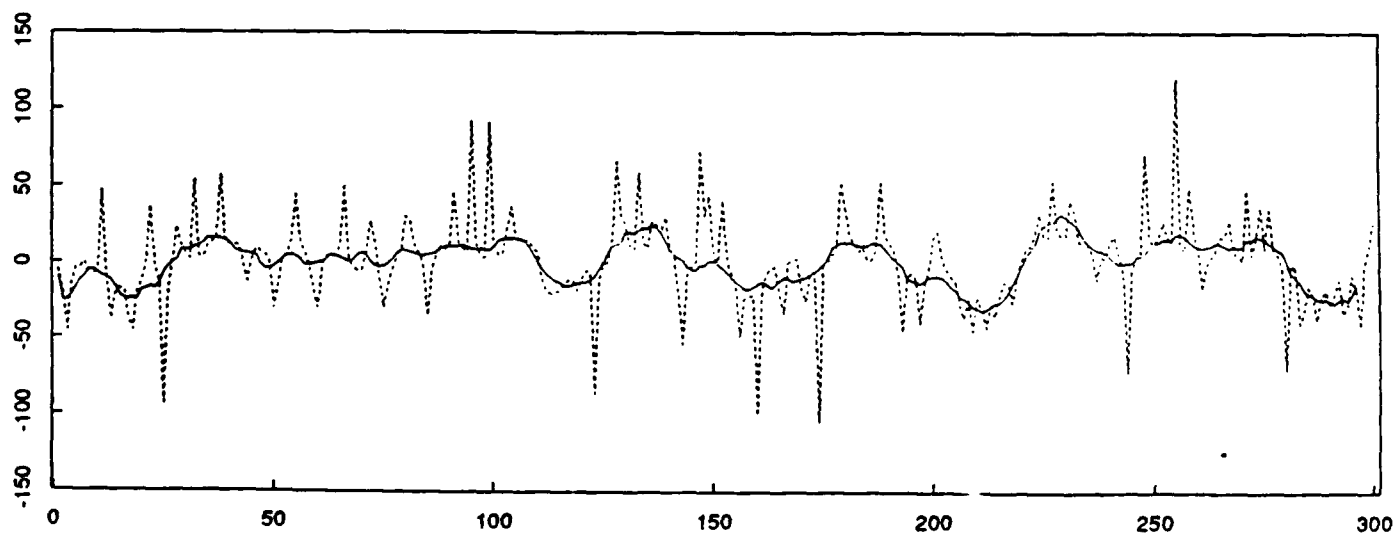
(c) THIRD ORDER DIFFERENCE MODEL, ESTIMATED INNOVATION VARIANCE = .04, OBSERVATION VARIANCE = 154.1

Figure 10: RADAR GLINT NOISE (---) AND SMOOTHERS (—) USING FIRST ORDER MODEL, ESTIMATED INNOVATION VARIANCE = 73.1, OBSERVATION VARIANCE = 57.2
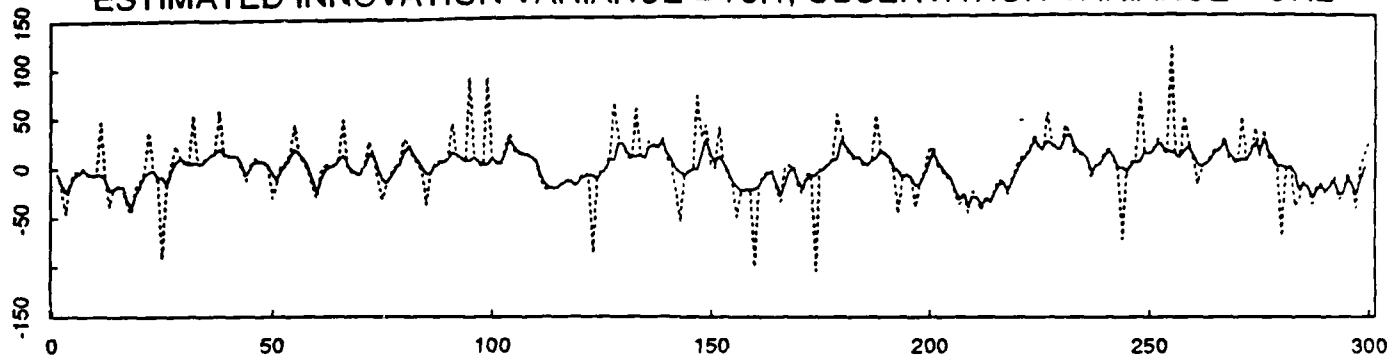
(a) LAG 0 SMOOTHER
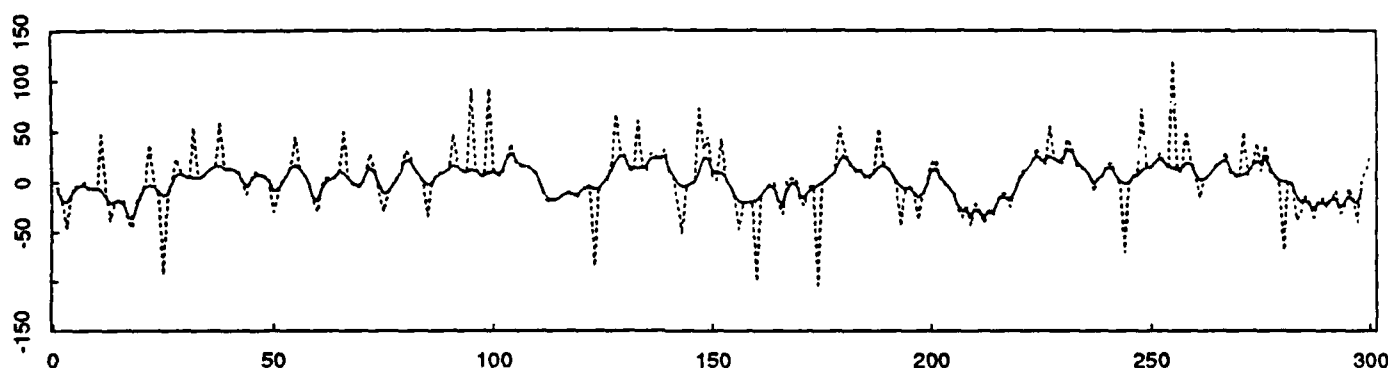
(b) LAG 1 SMOOTHER
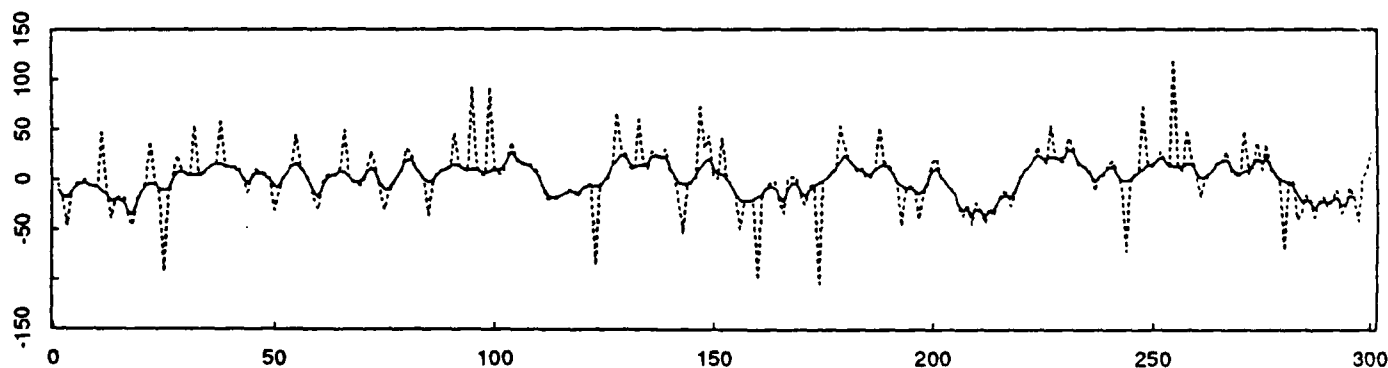
(c) LAG 3 SMOOTHER
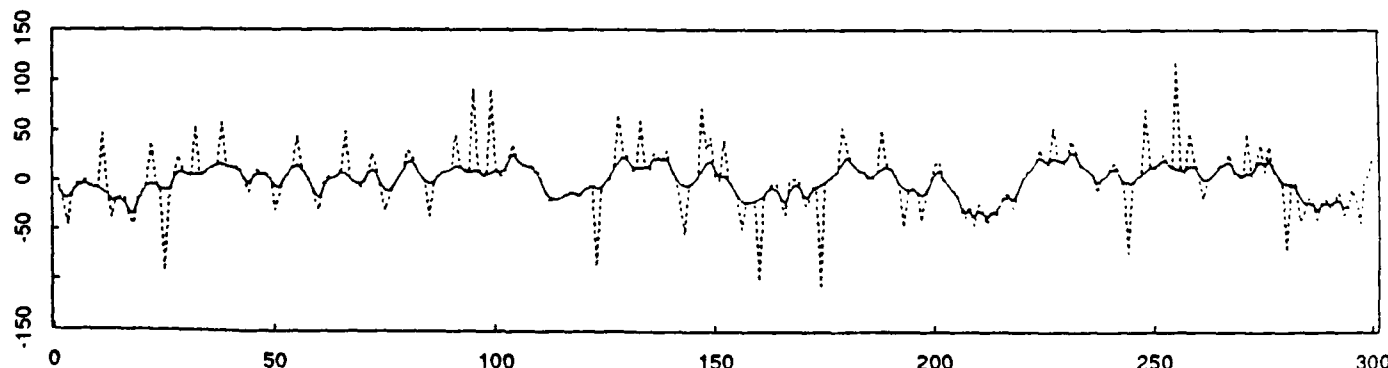
(d) LAG 5 SMOOTHER

Figure 11: RADAR GLINT NOISE (---) AND SMOOTHERS (—) USING SECOND ORDER MODEL, ESTIMATED INNOVATION VARIANCE = 1.62, OBSERVATION VARIANCE = 129.4
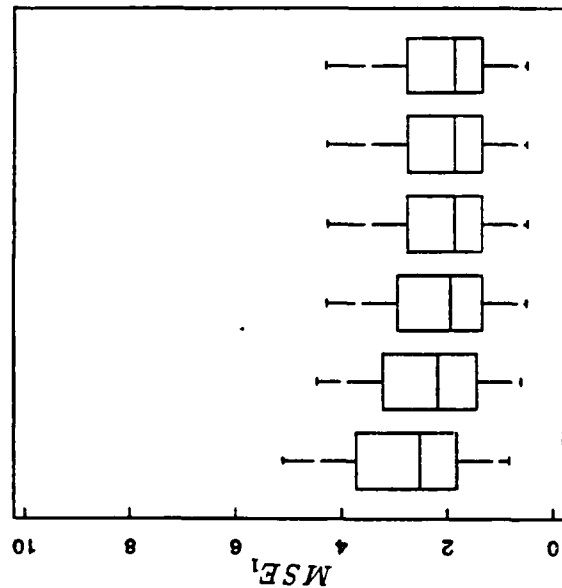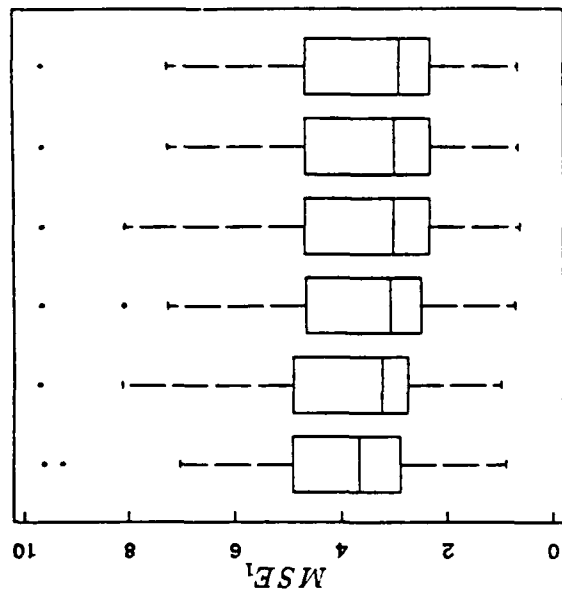
(a) LAG 0 SMOOTHER

(b) LAG 1 SMOOTHER

(c) LAG 3 SMOOTHER
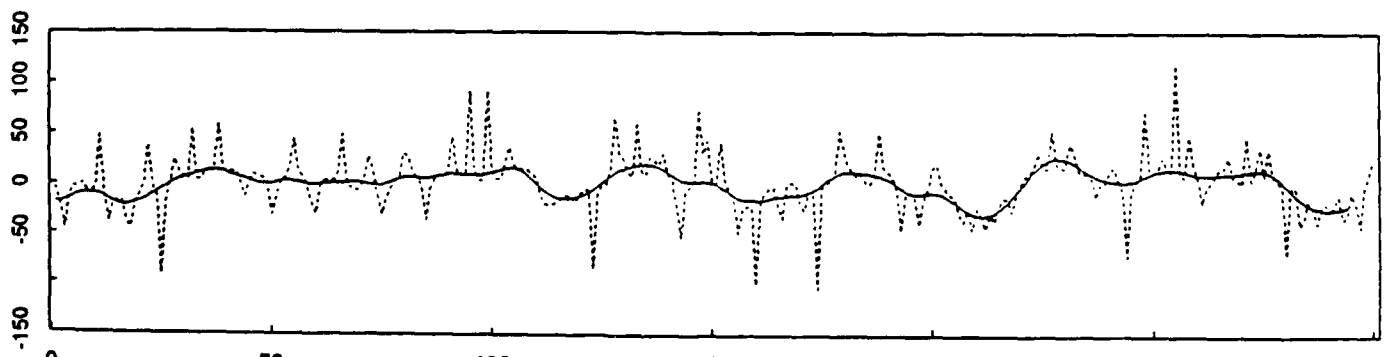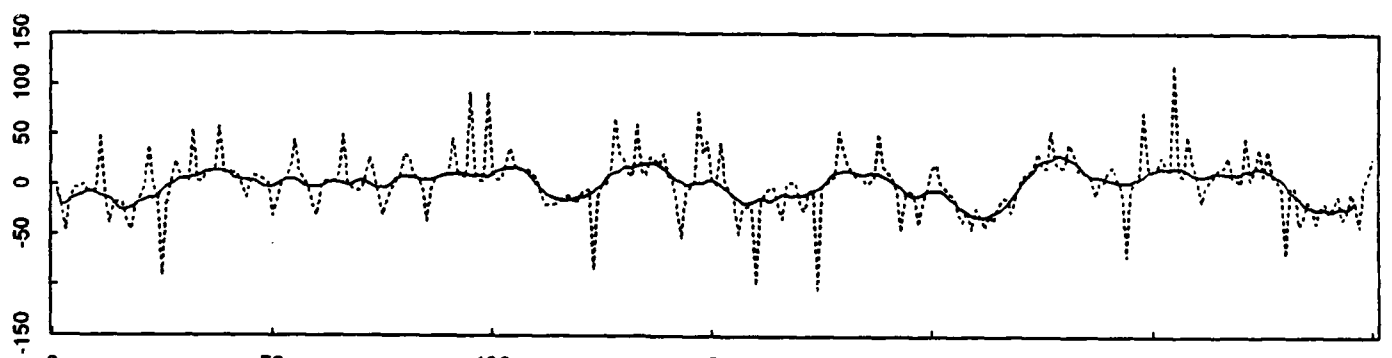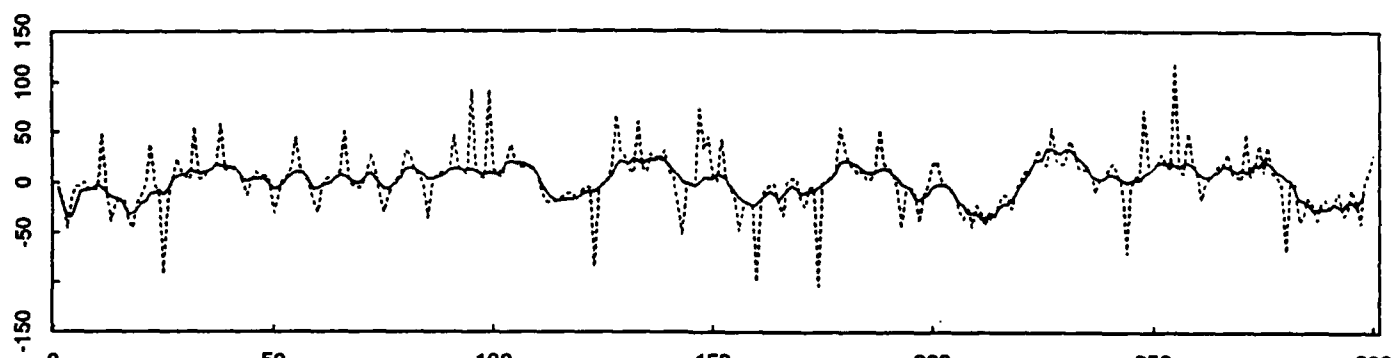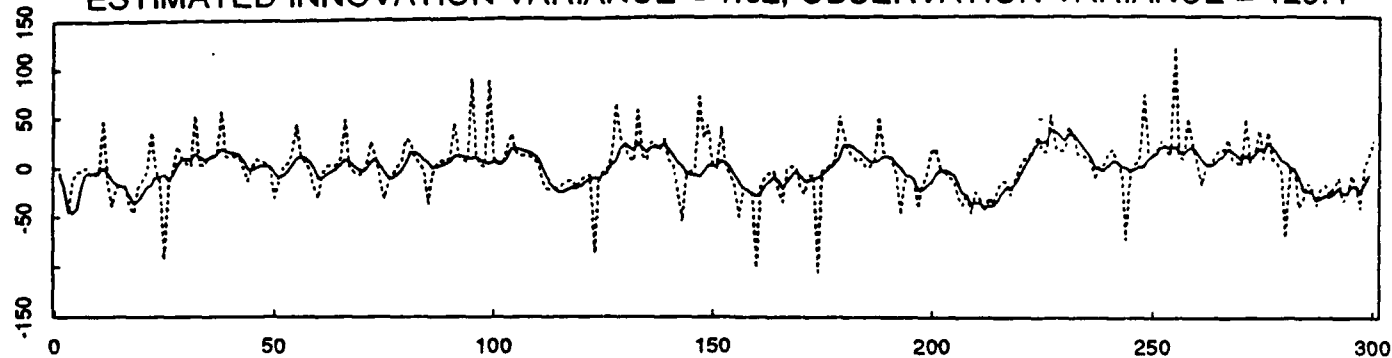
(d) LAG 5 SMOOTHER

Figure 12: MEAN SQUARED ERRORS AT OUTLIER (MSE_1) AND NON-OUTLIER (MSE_2) POSITIONS FOR LAG 0 TO LAG 5 AND OUTLIER PATCHES OF LENGTHS 1, 2 AND 3.
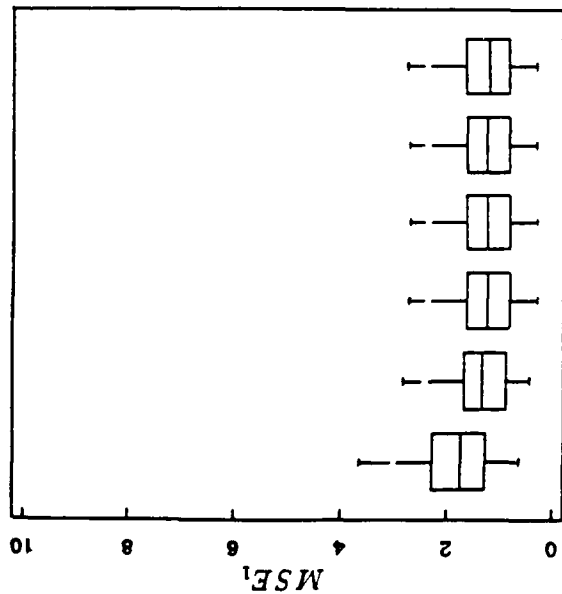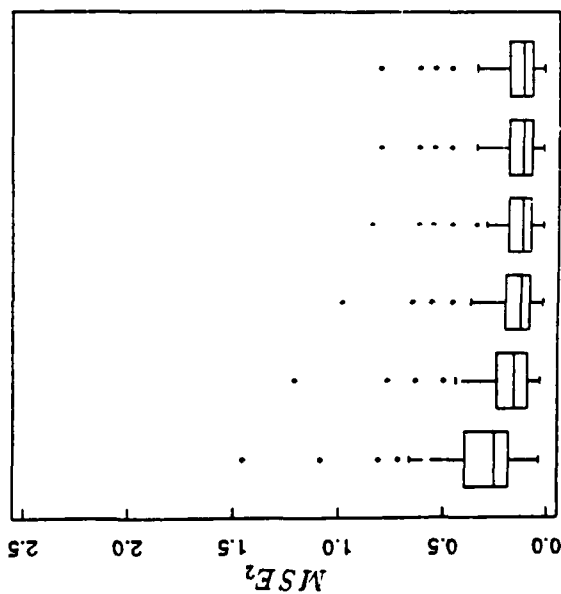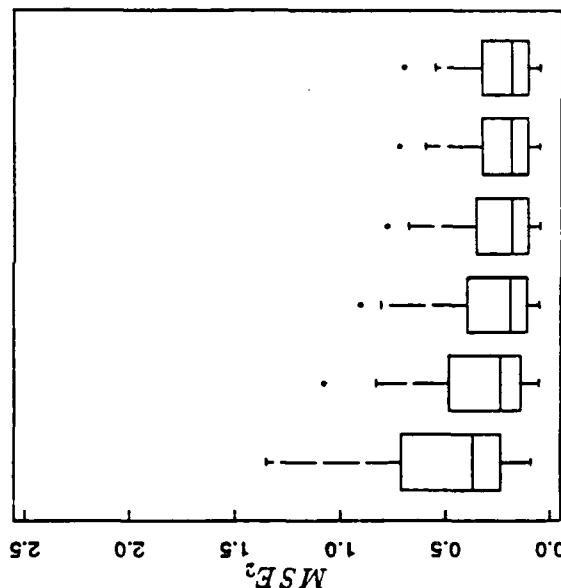
(a) PATCH = 1 AT OUTLIERS.
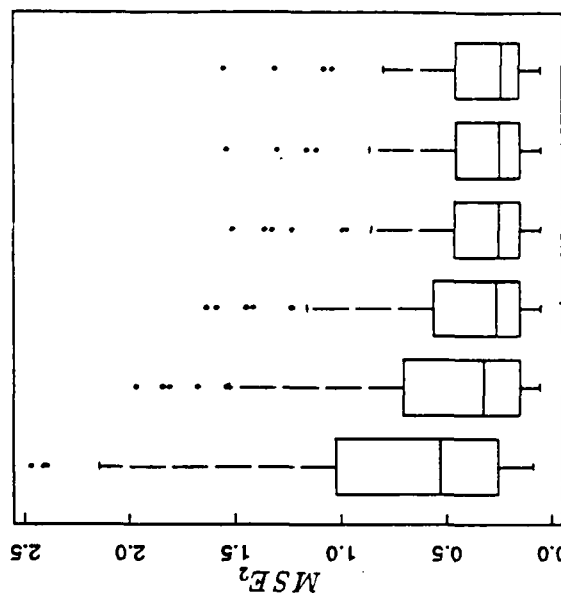
(b) PATCH = 2 AT OUTLIERS.

(c) PATCH = 3 AT OUTLIERS.

(d) PATCH = 1 AT NON-OUTLIERS.

(e) PATCH = 2 AT NON-OUTLIERS.

(f) PATCH = 3 AT NON-OUTLIERS.